

## B7-3 : Filtrage bayésien et approximation particulière

### TP 5 : Options européennes dans le modèle de Black-Scholes

suite du TP 3, en guise d'examen final

Une option européenne est définie comme le droit, qu'il n'est pas obligatoire d'exercer, d'acheter une quantité unitaire d'un actif à la date  $T$  et au prix  $K$ , la date d'exercice ou de maturité  $T$  et le prix d'exercice (strike)  $K$  étant fixés par contrat. La question se pose de déterminer le prix auquel une telle option doit être proposée. Dans le modèle de Black-Scholes, le prix unitaire  $S_t$  de l'actif à l'instant  $t$  est modélisé comme un mouvement brownien géométrique (exponentielle d'un mouvement brownien avec dérive)

$$dS_t = S_t (r dt + \sigma dW_t) ,$$

dont la solution explicite est donnée par la formule

$$S_t = S_0 \exp\left((r - \frac{1}{2}\sigma^2)t + \sigma W_t\right) .$$

La probabilité d'exercice est définie par

$$p = \mathbb{P}[S_T \geq K] = \mathbb{P}[S_0 \exp((r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}X) \geq K] ,$$

et on admettra que le prix de l'option à l'instant 0 est donné par la formule

$$P = e^{-rT} \mathbb{E}(S_T - K)^+ = e^{-rT} \mathbb{E}(S_0 \exp((r - \frac{1}{2}\sigma^2)T + \sigma\sqrt{T}X) - K)^+ ,$$

où  $X$  est une variable aléatoire gaussienne réduite centrée, de distribution de probabilité

$$\eta(dx) = \mathbb{P}[X \in dx] = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx .$$

Dans ce cas particulier très simple, il existe des expressions explicites

$$p = N\left(\frac{\log \frac{S_0}{K} + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) ,$$

et

$$P = S_0 N\left(\frac{\log \frac{S_0}{K} + (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) - e^{-rT} K N\left(\frac{\log \frac{S_0}{K} + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right) ,$$

où par définition

$$N(d) = \eta([-d, \infty)) = \frac{1}{\sqrt{2\pi}} \int_{-d}^{\infty} e^{-\frac{1}{2}x^2} dx ,$$

et on pourra utiliser la fonction MATLAB `erfc` pour calculer la valeur exacte du prix  $P$ . Par exemple, pour  $S_0 = 70$ ,  $K = 100$ ,  $T = 1$ ,  $r = 0$  et  $\sigma = 0.2$ , on a  $p = 0.0298$  et  $P = 0.2481$ .

L'objectif ici est de calculer numériquement la probabilité d'exercice  $p$  et le prix  $P$  de l'option en utilisant des méthodes de simulation Monte Carlo. En posant

$$V(x) = S_0 \exp\left(\left(r - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}x\right),$$

on voit qu'il s'agit, pour un niveau  $K$  donné, d'évaluer les quantités

$$p = \mathbb{P}[V(X) \geq K] \quad \text{et} \quad P_0 = e^{rT} P = \mathbb{E}(V(X) - K)^+.$$

Au vu des formules, on remarque qu'il s'agit d'un calcul d'intégrales dans la queue d'une distribution de probabilité gaussienne, et en introduisant une suite croissante de niveaux intermédiaires entre le prix  $S_0$  de l'actif à l'instant 0 et le prix d'exercice  $K$ , on a étudié dans le TP 3 une méthode de Monte Carlo avec interaction, pour

- estimer la probabilité d'exercice

$$p = \mathbb{P}[V(X) \geq K] = \int_{-\infty}^{\infty} g_K(x) \eta(dx),$$

- générer un échantillon distribué (approximativement) selon la loi conditionnelle

$$\mu_K(dx) = \mathbb{P}[X \in dx \mid V(X) \geq K] \quad \text{c'est-à-dire} \quad \mu_K = g_K \cdot \eta = \frac{g_K \eta}{\langle \eta, g_K \rangle},$$

- et finalement calculer (approximativement) le prix

$$P_0 = \mathbb{E}(V(X) - K)^+ = \int_{-\infty}^{\infty} (V(x) - K)^+ \eta(dx),$$

grâce à la représentation

$$P_0 = \int_{-\infty}^{\infty} (V(x) - K) g_K(x) \eta(dx) = p \int_{-\infty}^{\infty} (V(x) - K) \mu_K(dx),$$

où par définition  $g_K(x) = 1_{(V(x) \geq K)}$ .

La méthode traditionnelle par échantillonnage pondéré consiste à simuler un échantillon selon une autre distribution de probabilité, de telle sorte que le calcul d'intégrale concerne la zone centrale (et pas la queue) de la nouvelle distribution de probabilité, mais trouver la distribution de probabilité qui permet de construire un estimateur de variance minimale est en général difficile.

La méthode proposée dans le TP 3 est une alternative à la méthode traditionnelle par échantillonnage pondéré, et on se propose ici d'étudier une version améliorée, qui consiste à remarquer que l'échantillon obtenu à la dernière étape est distribué (approximativement) selon la distribution de probabilité optimale (qui minimise la variance de l'estimateur), et à utiliser cet échantillon pour mettre en oeuvre une méthode d'échantillonnage pondéré.

Soit à calculer une intégrale (ou une espérance mathématique) de la forme

$$I = \int_E g(x) \eta(dx) = \langle \eta, g \rangle ,$$

où  $g(x)$  est une fonction *positive* et  $\eta(dx)$  est une distribution de probabilité quelconque (pas nécessairement la distribution de probabilité gaussienne réduite centrée introduite plus haut). Une méthode traditionnelle pour calculer  $I$  est la méthode d'échantillonnage pondéré, où on utilise une autre distribution de probabilité  $\nu(dx)$ , telle que la dérivée  $r(x) = \frac{d\eta}{d\nu}(x)$  existe, de sorte que

$$I = \int_E g(x) \eta(dx) = \int_E g(x) r(x) \nu(dx) = \langle \nu, g r \rangle ,$$

et soit

$$I_M = \frac{1}{M} \sum_{j=1}^M g(x_j) r(x_j) ,$$

l'approximation Monte Carlo correspondante, où  $(x_1, \dots, x_M)$  est un  $M$ -échantillon distribué selon  $\nu(dx)$ .

- (i) Montrer que cet estimateur est sans biais, c'est-à-dire que  $\mathbb{E}(I_M) = I$ , avec une variance (non-asymptotique) donnée par

$$\mathbb{E}|I_M - I|^2 = \frac{1}{M} \text{var}(\nu, g r) \quad \text{et} \quad \text{var}(\nu, g r) = \int_E g^2(x) r(x) \eta(dx) - I^2 .$$

- (ii) Parmi tous les choix possibles pour l'estimation de  $I$ , montrer que la distribution de probabilité

$$\nu_* = g \cdot \eta = \frac{g \eta}{\langle \eta, g \rangle} ,$$

appelée *distribution d'importance optimale* pour l'estimation de  $I$ , conduit à un estimateur de variance nulle.

Même si la constante de normalisation  $\langle \eta, g \rangle$  n'est pas connue (sinon on connaîtrait  $I$  exactement), il est toujours possible en principe de simuler une variable aléatoire selon  $\nu_*(dx)$ , par exemple par une méthode d'acceptation / rejet. En revanche, la connaissance de la constante de normalisation est indispensable pour évaluer la dérivée  $r(x)$ , ce qui rend impraticable l'approximation Monte Carlo correspondant à la distribution d'importance optimale, et ce qui justifie l'intérêt pour des algorithmes adaptatifs, capables d'apprendre (approximativement) la distribution d'importance optimale, et qui conduisent à des estimateurs de variance petite, sinon nulle.

Par exemple, dans l'application aux options européennes, on a étudié dans le TP 3 une méthode de Monte Carlo avec interaction, capable de générer un échantillon distribué (approximativement) selon la loi conditionnelle  $\mu_K(dx)$ , qui est justement la distribution d'importance optimale pour le calcul de la probabilité d'exercice  $p$ .

(iii) Pour tout autre choix  $\nu(dx)$ , montrer que

$$\text{var}(\nu, g r) = \langle \eta, g \rangle^2 \chi^2(\nu_*, \nu) ,$$

en terme de la  $\chi^2$ -divergence

$$\chi^2(\nu_*, \nu) = \int_E (|\frac{d\nu_*}{d\nu}(x)|^2 - 1) \nu(dx) = \int_E (\frac{d\nu_*}{d\nu}(x) - 1)^2 \nu(dx) ,$$

entre  $\nu_*(dx)$  et  $\nu(dx)$  : en d'autres termes, utiliser une distribution d'importance  $\nu(dx)$  *proche* de la distribution d'importance optimale  $\nu_*(dx)$  au sens de la  $\chi^2$ -divergence, peut conduire à un estimateur de variance *petite*.

On considère la mesure positive (non-normalisée), et la distribution de probabilité associée, définies par

$$\langle \gamma, \phi \rangle = \int_E \phi(x) g(x) \eta(dx) = \langle \eta, g \phi \rangle \quad \text{et} \quad \langle \mu, \phi \rangle = \frac{\langle \gamma, \phi \rangle}{\langle \gamma, 1 \rangle} ,$$

respectivement.

(iv) Montrer que

- la constante de normalisation  $\langle \gamma, 1 \rangle$  fournit une autre expression de  $I$ ,
- la distribution normalisée  $\mu(dx)$  fournit une autre expression de la distribution d'importance optimale  $\nu_*(dx)$  pour l'estimation de  $I$ .

On suppose qu'il existe un noyau markovien  $M(x, dx')$ , pas nécessairement réversible, qui laisse invariante la distribution de probabilité  $\eta(dx)$ . Soit  $H_k(x)$  une fonction strictement positive pour tout  $k = 0, 1, \dots, n$ , avec en particulier  $H_n(x) = g(x)$  pour  $k = n$ .

(v) Montrer que

$$\langle \gamma, \phi \rangle = \int_E \dots \int_E \eta(dx_0) M(x_0, dx_1) \dots M(x_{k-1}, dx_k) \dots M(x_{n-1}, dx_n) g(x_n) \phi(x_n) ,$$

pour tout instant  $n$ , et que

$$\langle \gamma, \phi \rangle = \int_E \dots \int_E \gamma_0(dx_0) R_1(x_0, dx_1) \dots R_k(x_{k-1}, dx_k) \dots R_n(x_{n-1}, dx_n) \phi(x_n) ,$$

avec la mesure positive (non-normalisée)

$$\gamma_0(dx) = \eta(dx) H_0(x) ,$$

et avec le noyau positif (non-normalisé)

$$R_k(x, dx') = \frac{1}{H_{k-1}(x)} M(x, dx') H_k(x') ,$$

c'est-à-dire

$$R_k(x, dx') = W_k(x, x') M(x, dx') \quad \text{avec} \quad W_k(x, x') = \frac{H_k(x')}{H_{k-1}(x)} ,$$

pour tout  $k = 1, \dots, n$ .

Au vu du résultat obtenu en réponse à la question (v), on introduit pour tout  $k = 1, \dots, n$  la mesure positive (non-normalisée), et la distribution de probabilité associée, définies par

$$\langle \gamma_k, \phi \rangle = \int_E \cdots \int_E \gamma_0(dx_0) R_1(x_0, dx_1) \cdots R_k(x_{k-1}, dx_k) \phi(x_k) \quad \text{et} \quad \langle \mu_k, \phi \rangle = \frac{\langle \gamma_k, \phi \rangle}{\langle \gamma_k, 1 \rangle},$$

respectivement, et on remarque que

$$\mu R_k(dx') = \int_E \mu(dx) R_k(x, dx') = \int_E W_k(x, x') \mu(dx) M(x, dx'),$$

de sorte que la mesure positive  $\mu R_k(dx')$  apparaît comme la mesure marginale (obtenue en intégrant par rapport à la variable  $x$ ) de la mesure positive

$$(W_k(\mu \otimes M))(dx, dx') = W_k(x, x') (\mu \otimes M)(dx, dx') = W_k(x, x') \mu(dx) M(x, dx'),$$

définie sur l'ensemble produit  $E \times E$ .

- (vi) En déduire, en utilisant les résultats du cours, l'expression de la distribution non-normalisée  $\gamma_k(dx)$  en fonction de la distribution non-normalisée  $\gamma_{k-1}(dx)$ , puis en fonction de la distribution normalisée  $\mu_{k-1}(dx)$  et de la constante de normalisation  $\langle \gamma_{k-1}, 1 \rangle$ .

On recherche une approximation particulière pondérée de la forme

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i \delta_{\xi_k^i} \quad \text{avec} \quad \sum_{i=1}^N w_k^i = 1.$$

- (vii) Décrire, en utilisant les résultats du cours, l'approximation particulière de type bootstrap pour la constante de normalisation  $\langle \gamma_k, 1 \rangle$  et la distribution normalisée  $\mu_k(dx)$  : montrer en particulier que

$$\langle \gamma_0, 1 \rangle \approx \langle \gamma_0^N, 1 \rangle = \sum_{i=1}^N H_0(\xi_0^i),$$

et

$$\mu_0 \approx \mu_0^N = \sum_{i=1}^N w_0^i \delta_{\xi_0^i} \quad \text{avec} \quad w_0^i \propto H_0(\xi_0^i),$$

où indépendamment pour tout  $i = 1, \dots, N$ , on simule  $\xi_0^i$  selon  $\eta(dx)$ , et récursivement pour tout  $k = 1, \dots, n$

$$\langle \gamma_k, 1 \rangle \approx \langle \gamma_k^N, 1 \rangle = \left[ \frac{1}{N} \sum_{i=1}^N \frac{H_k(\xi_k^i)}{H_{k-1}(\xi_{k-1}^{\tau_k^i})} \right] \langle \gamma_{k-1}^N, 1 \rangle,$$

et

$$\mu_k \approx \mu_k^N = \sum_{i=1}^N w_k^i \delta_{\xi_k^i} \quad \text{avec} \quad w_k^i \propto W_k(\xi_{k-1}^{\tau_k^i}, \xi_k^i) = \frac{H_k(\xi_k^i)}{H_{k-1}(\xi_{k-1}^{\tau_k^i})},$$

où indépendamment pour tout  $i = 1, \dots, N$ , on simule  $\tau_k^i$  à valeurs dans l'ensemble  $\{1, \dots, N\}$  des indices selon le vecteur de probabilités  $(w_{k-1}^1, \dots, w_{k-1}^N)$  et on simule  $\xi_k^i$  selon  $M(\xi_{k-1}^{\tau_k^i}, dx')$ , ce qui garantit que  $(\xi_{k-1}^{\tau_k^i}, \xi_k^i)$  est distribué selon  $(\mu_{k-1}^N \otimes M)(dx, dx')$ .

En particulier pour  $k = n$ , et au vu de la réponse à la question (iv), on obtient une première approximation de  $I$ , paramétrée par le nombre  $N$  de particules, à savoir

$$I_N = \langle \gamma_n^N, 1 \rangle ,$$

et une première approximation de la distribution d'importance optimale pour l'estimation de  $I$ , paramétrée par le nombre  $N$  de particules, à savoir

$$\nu_N = \hat{\mu}_n^N = \sum_{i=1}^N w_n^i \delta_{\xi_n^i} .$$

On suppose à partir de maintenant que  $E = \mathbb{R}^m$ , et que la distribution de probabilité  $\eta(dx) = p(x) dx$  possède une densité  $p(x)$ . On introduit une famille de noyaux régularisants : à partir d'une densité de probabilité  $K(u)$  définie sur  $\mathbb{R}^m$ , on pose  $K_h(u) = h^{-m} K(h^{-1} u)$  pour tout  $h > 0$ . On construit ainsi par convolution avec le noyau régularisant, une seconde approximation de la distribution d'importance optimale pour l'estimation de  $I$ , paramétrée par le nombre  $N$  de particules et par la largeur  $h$  de la fenêtre, c'est-à-dire

$$\nu_{N,h}(dx) = (\mu_n^N * K_h)(dx) = \underbrace{\sum_{i=1}^N w_n^i K_h(x - \xi_n^i)}_{q_{N,h}(x)} dx ,$$

et en utilisant cette approximation comme distribution d'importance, on obtient une seconde approximation de  $I$ , paramétrée par le nombre  $N$  de particules, par la taille  $M$  de l'échantillon et par la largeur  $h$  de la fenêtre, c'est-à-dire

$$I_{N,M,h} = \frac{1}{M} \sum_{j=1}^M g(x_j) r(x_j) \quad \text{avec} \quad r(x) = \frac{d\eta}{d\nu_{N,h}}(x) = \frac{p(x)}{q_{N,h}(x)} ,$$

où  $(x_1, \dots, x_M)$  est un  $M$ -échantillon distribué selon  $\nu_{N,h}(dx) = q_{N,h}(x) dx$ .

## APPLICATION AUX OPTIONS EUROPÉENNES

On rappelle qu'il s'agit de calculer

$$p = \int_{-\infty}^{\infty} 1_{(V(x) \geq K)} \eta(dx) \quad \text{et} \quad P_0 = \int_{-\infty}^{\infty} \max(V(x) - K, 0) \eta(dx) ,$$

pour la fonction

$$V(x) = S_0 \exp\left(\left(r - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}x\right) ,$$

et pour la distribution de probabilité

$$\eta(dx) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx .$$

On rappelle également que le noyau markovien

$$M(x, dx') = \frac{1}{\sqrt{2\pi(1-c^2)}} \exp\left(-\frac{1}{2} \frac{(x' - cx)^2}{1-c^2}\right) dx' , \quad \text{avec } |c| < 1,$$

est réversible pour la distribution de probabilité  $\eta(dx)$ , et donc a fortiori laisse  $\eta(dx)$  invariante, et que pour simuler une variable aléatoire  $\Xi$  distribuée selon  $M(x, dx')$ , où  $x \in E$  est fixé, il suffit de simuler une variable aléatoire gaussienne  $W$  centrée réduite, et de poser  $\Xi = cx + \sqrt{1-c^2} W$ .

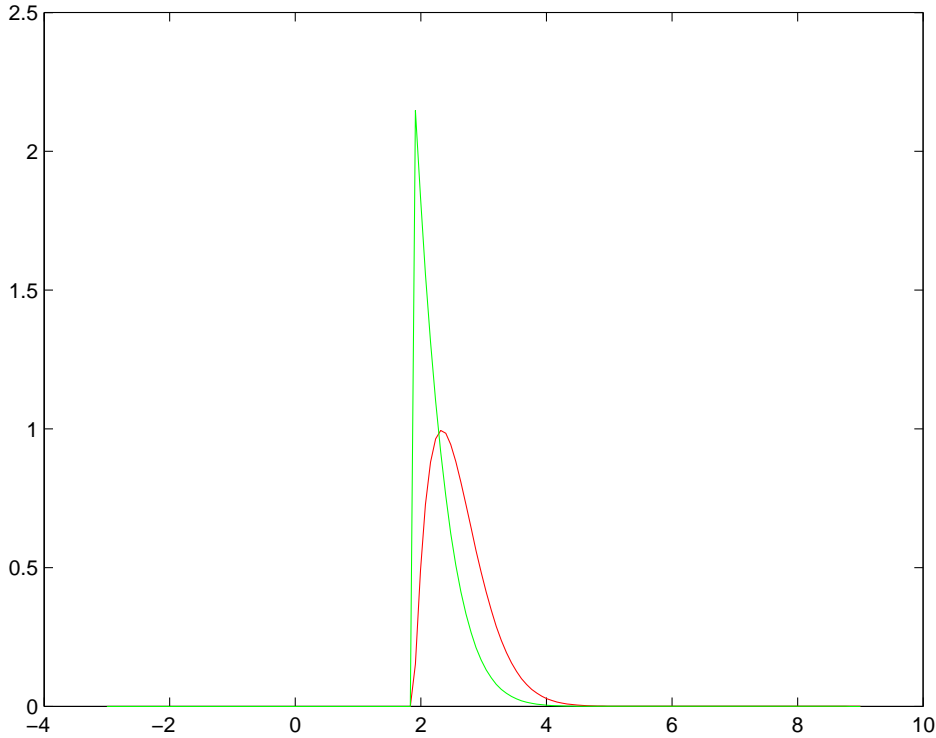


Figure 1: Densités d'importance optimales

**Estimation de la probabilité d'exercice  $p$  :** La distribution d'importance

$$\mu_K = g_K \cdot \eta = \frac{g_K \eta}{\langle \eta, g_K \rangle} \quad \text{avec} \quad g_K(x) = 1_{(V(x) \geq K)} ,$$

représentée Figure 1 (en vert), est optimale pour l'estimation de la probabilité d'exercice  $p$ . On suit l'approche décrite à la question (v) pour l'apprentissage de cette distribution d'importance optimale : on pose  $H_n(x) = 1_{(V(x) \geq K)}$ , et en introduisant une suite croissante

$$S_0 = a_0 < \dots < a_k < \dots < a_n = K ,$$

de niveaux entre le prix  $S_0$  de l'actif à l'instant 0 et le prix d'exercice  $K$ , on définit

$$H_k(x) = \varepsilon \cdot 1_{(V(x) < a_k)} + 1_{(V(x) \geq a_k)} = \varepsilon + (1 - \varepsilon) \cdot 1_{(V(x) \geq a_k)},$$

pour tout  $k = 0, 1, \dots, (n - 1)$ , avec  $\varepsilon > 0$ .

- (viii) Mettre en œuvre l'algorithme particulière décrit à la question (vii) pour estimer la probabilité d'exercice  $p$  et pour générer un  $N$ -échantillon distribué (approximativement) selon la distribution d'importance optimale  $\mu_K(dx)$ . Avec l'échantillon obtenu, calculer une première approximation du prix  $P_0$ . On utilisera des niveaux régulièrement espacés entre le prix  $S_0$  de l'actif à l'instant 0 et le prix d'exercice  $K$ , et on prendra les valeurs numériques  $c = 0.9$  pour le noyau réversible et  $\varepsilon = 0.01$  pour les approximations successives de la fonction  $g_K(x)$ . On pourra par exemple représenter à l'aide d'un histogramme l'approximation particulière obtenue à la dernière étape, comme à la Figure 2 (en jaune).

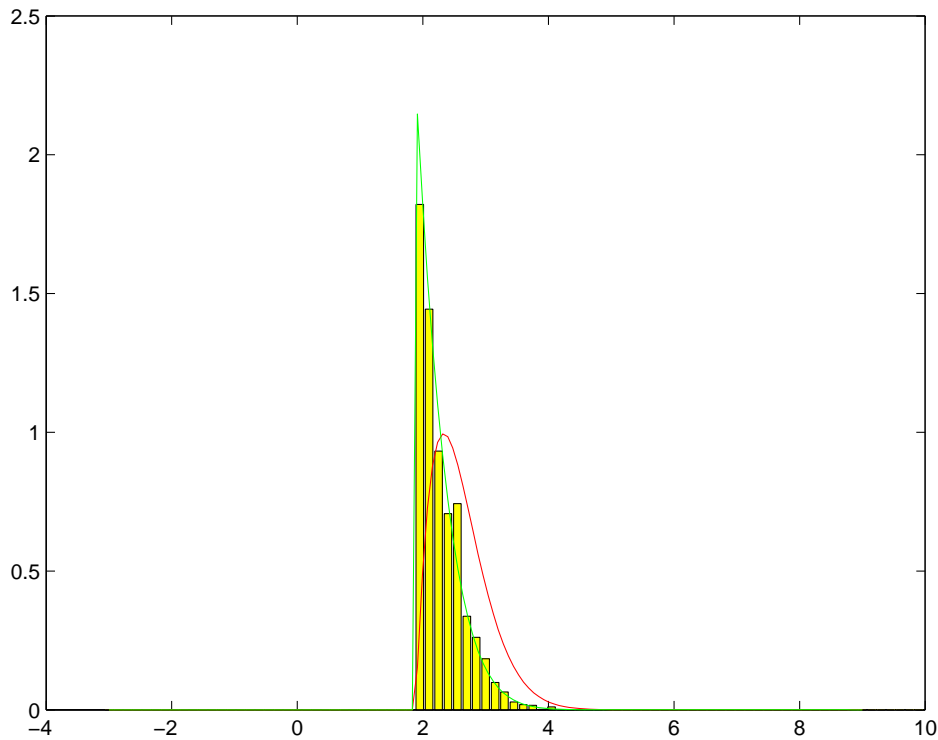


Figure 2: Apprentissage de la distribution d'importance optimale pour l'estimation de la probabilité d'exercice  $p$

- (ix) Construire par convolution avec un noyau régularisant, une seconde approximation de la distribution d'importance optimale  $\mu_K(dx)$  et générer un  $M$ -échantillon distribué selon

l'approximation régularisée. Avec l'échantillon obtenu, calculer une seconde approximation de la probabilité d'exercice  $p$  et du prix  $P_0$ . On utilisera une densité gaussienne réduite centrée pour la famille de noyaux régularisants, et on prendra la valeur numérique  $h = 0.1$  pour la largeur de la fenêtre. On représentera la densité obtenue par régularisation, comme à la Figure 3 (en bleu).

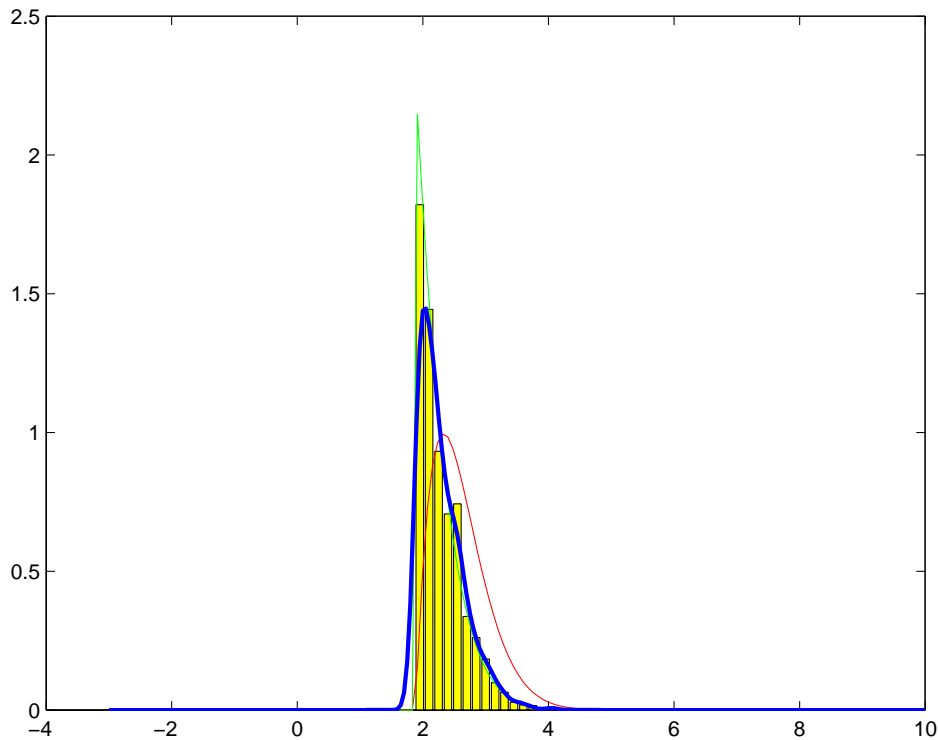


Figure 3: Approximation régularisée de la densité d'importance optimale pour l'estimation de la probabilité d'exercice  $p$

**Estimation du prix  $P_0$  :** La distribution d'importance  $\mu_K(dx)$  est optimale pour l'estimation de la probabilité d'exercice  $p$ , mais pas pour l'estimation du prix  $P_0$  : en revanche, la distribution d'importance

$$\nu_K = h_K \cdot \eta = \frac{h_K \eta}{\langle \eta, h_K \rangle} \quad \text{avec} \quad h_K(x) = \max(V(x) - K, 0) ,$$

représentée Figure 1 (en rouge), est optimale pour l'estimation du prix  $P_0$ . On suit l'approche décrite à la question (v) pour l'apprentissage de cette distribution d'importance optimale : on

pose  $H_n(x) = \max(V(x) - K, 0)$ , et en introduisant une suite croissante

$$S_0 = a_0 < \dots < a_k < \dots < a_n = K ,$$

de niveaux entre le prix  $S_0$  de l'actif à l'instant 0 et le prix d'exercice  $K$ , on définit

$$H_k(x) = \varepsilon \cdot 1_{(V(x) < a_k)} + \max(V(x) - a_k, 0) = \varepsilon + (V(x) - a_k - \varepsilon) \cdot 1_{(V(x) \geq a_k)} ,$$

ou bien alternativement

$$H_k(x) = \max(V(x) - a_k, \varepsilon) = \varepsilon + (V(x) - a_k - \varepsilon) \cdot 1_{(V(x) \geq a_k + \varepsilon)} ,$$

pour tout  $k = 0, 1, \dots, (n - 1)$ , avec  $\varepsilon > 0$ .

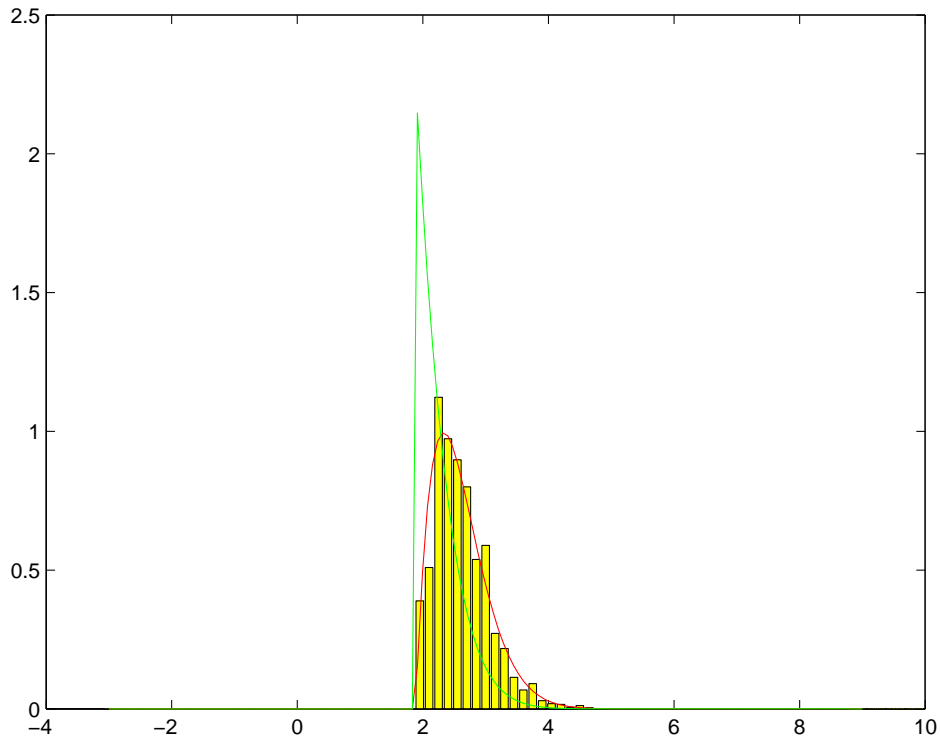


Figure 4: Apprentissage de la distribution d'importance optimale pour l'estimation du prix  $P_0$

- (x) Mettre en œuvre l'algorithme particulier décrit à la question (vii) pour estimer le prix  $P_0$  et pour générer un  $N$ -échantillon distribué (approximativement) selon la distribution d'importance optimale  $\nu_K(dx)$ . On utilisera des niveaux régulièrement espacés entre le prix  $S_0$  de l'actif à l'instant 0 et le prix d'exercice  $K$ , et on prendra les valeurs numériques  $c =$

0.9 pour le noyau réversible et  $\varepsilon = 0.1$  pour les approximations successives de la fonction  $h_K(x)$ . On pourra par exemple représenter à l'aide d'un histogramme l'approximation particulière obtenue à la dernière étape, comme à la Figure 4 (en jaune).

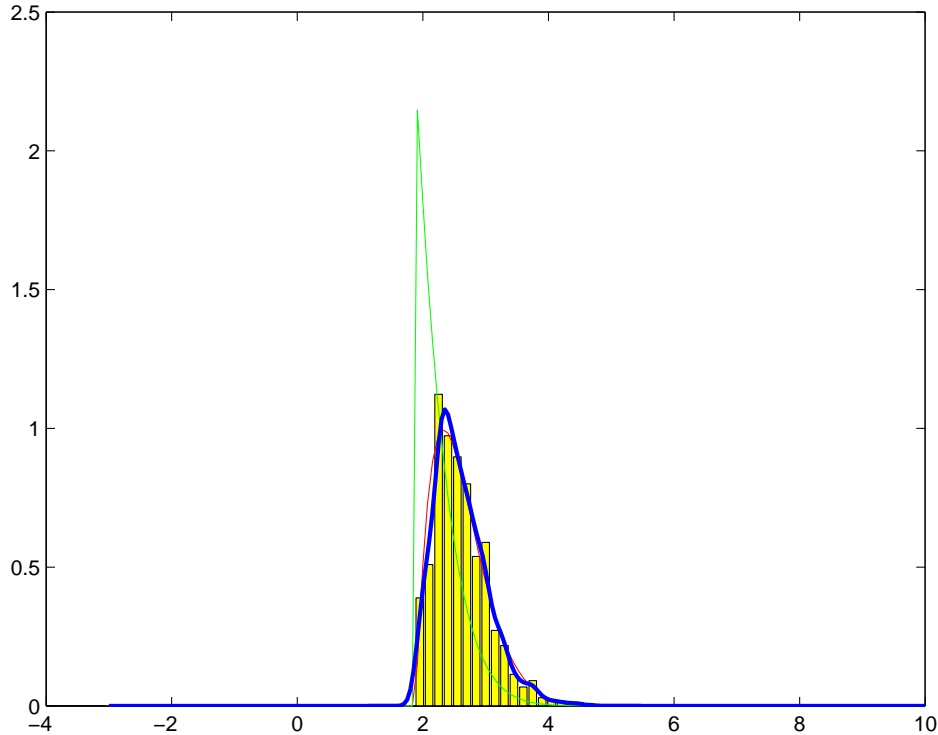


Figure 5: Approximation régularisée de la densité d'importance optimale pour l'estimation du prix  $P_0$

- (xi) Construire par convolution avec un noyau régularisant, une seconde approximation de la distribution d'importance optimale  $\nu_K(dx)$  et générer un  $M$ -échantillon distribué selon l'approximation régularisée. Avec l'échantillon obtenu, calculer une seconde approximation du prix  $P_0$ . On utilisera une densité gaussienne réduite centrée pour la famille de noyaux régularisants, et on prendra la valeur numérique  $h = 0.1$  pour la largeur de la fenêtre. On représentera la densité obtenue par régularisation, comme à la Figure 5 (en bleu).

On s'intéressera à la variance des estimateurs, dont une valeur approchée sera obtenue en répétant plusieurs simulations Monte Carlo indépendantes, et on vérifiera la réduction de variance apportée par le post-traitement mis en œuvre aux questions (ix) et (xi), par rapport à la méthode mise en œuvre aux questions (viii) et (x) ou lors du TP 3.