Échantillonnage stratégique (ou pondéré) "importance sampling (IS)"

On remplace la densité π d'un vecteur de v.a.'s $\mathbf{Y}$ par une autre densité, g, telle que $g(\mathbf{y}) > 0$ lorsque $h(\mathbf{y})\pi(\mathbf{y}) \neq 0$. On a

$$\mu = \mathbb{E}_{\pi}[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y})\pi(\mathbf{y})d\mathbf{y}$$

Échantillonnage stratégique (ou pondéré) “importance sampling (IS)”

On remplace la densité π d'un vecteur de v.a.'s $\mathbf{Y}$ par une autre densité, g, telle que $g(\mathbf{y}) > 0$ lorsque $h(\mathbf{y})\pi(\mathbf{y}) \neq 0$. On a

$$\mu \quad = \quad \mathbb{E}_\pi[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y})\pi(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^d} [h(\mathbf{y})\pi(\mathbf{y})/g(\mathbf{y})]g(\mathbf{y})d\mathbf{y}$$

Échantillonnage stratégique (ou pondéré) "importance sampling (IS)"

On remplace la densité π d'un vecteur de v.a.'s $\mathbf{Y}$ par une autre densité, g, telle que $g(\mathbf{y}) > 0$ lorsque $h(\mathbf{y})\pi(\mathbf{y}) \neq 0$. On a

$$\begin{aligned} \mu &= \mathbb{E}_\pi[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y})\pi(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^d} [h(\mathbf{y})\pi(\mathbf{y})/g(\mathbf{y})]g(\mathbf{y})d\mathbf{y} \\ &= \mathbb{E}_g[h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})]. \end{aligned}$$

Échantillonnage stratégique (ou pondéré) "importance sampling (IS)"

On remplace la densité π d'un vecteur de v.a.'s $\mathbf{Y}$ par une autre densité, g, telle que $g(\mathbf{y}) > 0$ lorsque $h(\mathbf{y})\pi(\mathbf{y}) \neq 0$. On a

$$\begin{aligned}
\mu &= \mathbb{E}_{\pi}[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y})\pi(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^d} [h(\mathbf{y})\pi(\mathbf{y})/g(\mathbf{y})]g(\mathbf{y})d\mathbf{y} \\
&= \mathbb{E}_{g}[h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})].
\end{aligned}$$

On génère $\mathbf{Y}$ selon g, et l'estimateur est $X_{\text{is}} = h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})$.

Échantillonnage stratégique (ou pondéré) "importance sampling (IS)"

On remplace la densité π d'un vecteur de v.a.'s $\mathbf{Y}$ par une autre densité, g, telle que $g(\mathbf{y}) > 0$ lorsque $h(\mathbf{y})\pi(\mathbf{y}) \neq 0$. On a

$$\begin{aligned}
\mu &= \mathbb{E}_\pi[h(\mathbf{Y})] = \int_{\mathbb{R}^d} h(\mathbf{y})\pi(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^d} [h(\mathbf{y})\pi(\mathbf{y})/g(\mathbf{y})]g(\mathbf{y})d\mathbf{y} \\
&= \mathbb{E}_g[h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})].
\end{aligned}$$

On génère $\mathbf{Y}$ selon g, et l'estimateur est $X_{\text{is}} = h(\mathbf{Y})\pi(\mathbf{Y})/g(\mathbf{Y})$.

Classe importante d'applications: simulation d'événements rares.

Cas discret. $\mathbb{P}\{Y = y_k\} = p(y_k)$ pour $k = 0, 1, \ldots$.

Cas discret. $\mathbb{P}\{Y = y_k\} = p(y_k)$ pour $k = 0, 1, \ldots$.

On change les probabilités $p(y_k)$ pour des probabilités $q(y_k)$ telles que $q(y_k) > 0$ lorsque $h(y_k)p(y_k) \neq 0$. Alors

$$\mu = \mathbb{E}_p[h(Y)] = \sum_{k=0}^{\infty} h(y_k)p(y_k)$$

Cas discret. $\mathbb{P}\{Y = y_k\} = p(y_k)$ pour $k = 0, 1, \ldots$.

On change les probabilités $p(y_k)$ pour des probabilités $q(y_k)$ telles que $q(y_k) > 0$ lorsque $h(y_k)p(y_k) \neq 0$. Alors

$$\mu = \mathbb{E}_p[h(Y)] = \sum_{k=0}^{\infty} h(y_k)p(y_k) = \sum_{k=0}^{\infty} [h(y_k)p(y_k)/q(y_k)]q(y_k)$$

Cas discret. $\mathbb{P}\{Y = y_k\} = p(y_k)$ pour $k = 0, 1, \ldots$.

On change les probabilités $p(y_k)$ pour des probabilités $q(y_k)$ telles que $q(y_k) > 0$ lorsque $h(y_k)p(y_k) \neq 0$. Alors

$$\begin{aligned}
\mu &= \mathbb{E}_p[h(Y)] = \sum_{k=0}^{\infty} h(y_k)p(y_k) = \sum_{k=0}^{\infty} [h(y_k)p(y_k)/q(y_k)]q(y_k) \\
&= \mathbb{E}_q[h(Y)p(Y)/q(Y)]
\end{aligned}$$

Cas discret. $\mathbb{P}\{Y = y_k\} = p(y_k)$ pour $k = 0, 1, \ldots$.

On change les probabilités $p(y_k)$ pour des probabilités $q(y_k)$ telles que $q(y_k) > 0$ lorsque $h(y_k)p(y_k) \neq 0$. Alors

$$\begin{aligned} \mu &= \mathbb{E}_p[h(Y)] = \sum_{k=0}^{\infty} h(y_k)p(y_k) = \sum_{k=0}^{\infty} [h(y_k)p(y_k)/q(y_k)]q(y_k) \\ &= \mathbb{E}_q[h(Y)p(Y)/q(Y)] \end{aligned}$$

Estimateur sans biais de μ: $X_{\text{is}} = h(Y)p(Y)/q(Y)$.

Supposons que l'on a $X = h(Y_1, \ldots, Y_d)$ où $\mathbf{Y} = (Y_1, \ldots, Y_d)$ sont des v.a. indépendantes de densités $\pi_1, \ldots, \pi_d$.

Supposons que l'on a $X = h(Y_1, \ldots, Y_d)$ où $\mathbf{Y} = (Y_1, \ldots, Y_d)$ sont des v.a. indépendantes de densités $\pi_1, \ldots, \pi_d$.

Une façon simple d'appliquer IS est de remplacer chaque π_j par une autre densité g_j. Les Y_j demeurent indépendantes.

Supposons que l'on a $X = h(Y_1, \ldots, Y_d)$ où $\mathbf{Y} = (Y_1, \ldots, Y_d)$ sont des v.a. indépendantes de densités $\pi_1, \ldots, \pi_d$.

Une façon simple d'appliquer IS est de remplacer chaque π_j par une autre densité g_j. Les Y_j demeurent indépendantes. On a

$$\mu = \mathbb{E}_\pi[h(Y_1, \ldots, Y_d)]$$

Supposons que l'on a $X = h(Y_1, \ldots, Y_d)$ où $\mathbf{Y} = (Y_1, \ldots, Y_d)$ sont des v.a. indépendantes de densités $\pi_1, \ldots, \pi_d$.

Une façon simple d'appliquer IS est de remplacer chaque π_j par une autre densité g_j. Les Y_j demeurent indépendantes. On a

$$\begin{aligned} \mu &= \mathbb{E}_\pi[h(Y_1, \ldots, Y_d)] \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \ldots, y_d)\pi_1(y_1) \cdots \pi_d(y_d) dy_1 \cdots dy_d \end{aligned}$$

Supposons que l'on a $X = h(Y_1, \ldots, Y_d)$ où $\mathbf{Y} = (Y_1, \ldots, Y_d)$ sont des v.a. indépendantes de densités $\pi_1, \ldots, \pi_d$.

Une façon simple d'appliquer IS est de remplacer chaque π_j par une autre densité g_j. Les Y_j demeurent indépendantes. On a

$$
\begin{aligned}
\mu &= \mathbb{E}_\pi[h(Y_1, \ldots, Y_d)] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \ldots, y_d) \pi_1(y_1) \cdots \pi_d(y_d) dy_1 \cdots dy_d \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[h(y_1, \ldots, y_d) \frac{\pi_1(y_1) \cdots \pi_d(y_d)}{g_1(y_1) \cdots g_d(y_d)} \right] g_1(y_1) \cdots g_d(y_d) dy_1 \cdots dy_d
\end{aligned}
$$

Supposons que l'on a $X = h(Y_1, \ldots, Y_d)$ où $\mathbf{Y} = (Y_1, \ldots, Y_d)$ sont des v.a. indépendantes de densités $\pi_1, \ldots, \pi_d$.

Une façon simple d'appliquer IS est de remplacer chaque π_j par une autre densité g_j. Les Y_j demeurent indépendantes. On a

$$
\begin{aligned}
\mu &= \mathbb{E}_\pi[h(Y_1, \ldots, Y_d)] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \ldots, y_d) \pi_1(y_1) \cdots \pi_d(y_d) dy_1 \cdots dy_d \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[h(y_1, \ldots, y_d) \frac{\pi_1(y_1) \cdots \pi_d(y_d)}{g_1(y_1) \cdots g_d(y_d)} \right] g_1(y_1) \cdots g_d(y_d) dy_1 \cdots dy_d \\
&= \mathbb{E}_g \left[h(Y_1, \ldots, Y_d) \frac{\pi_1(Y_1) \cdots \pi_d(Y_d)}{g_1(Y_1) \cdots g_d(Y_d)} \right],
\end{aligned}
$$

en supposant que $g_1(y_1) \cdots g_d(y_d) > 0$ lorsque $h(y_1, \ldots, y_d) \pi_1(y_1) \cdots \pi_d(y_d) \neq 0$.

Supposons que l'on a $X = h(Y_1, \ldots, Y_d)$ où $\mathbf{Y} = (Y_1, \ldots, Y_d)$ sont des v.a. indépendantes de densités $\pi_1, \ldots, \pi_d$.

Une façon simple d'appliquer IS est de remplacer chaque π_j par une autre densité g_j. Les Y_j demeurent indépendantes. On a

$$\begin{aligned}
\mu &= \mathbb{E}_\pi[h(Y_1, \ldots, Y_d)] \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(y_1, \ldots, y_d)\pi_1(y_1) \cdots \pi_d(y_d) dy_1 \cdots dy_d \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[h(y_1, \ldots, y_d) \frac{\pi_1(y_1) \cdots \pi_d(y_d)}{g_1(y_1) \cdots g_d(y_d)} \right] g_1(y_1) \cdots g_d(y_d) dy_1 \cdots dy_d \\
&= \mathbb{E}_g \left[h(Y_1, \ldots, Y_d) \frac{\pi_1(Y_1) \cdots \pi_d(Y_d)}{g_1(Y_1) \cdots g_d(Y_d)} \right],
\end{aligned}$$

en supposant que $g_1(y_1) \cdots g_d(y_d) > 0$ lorsque $h(y_1, \ldots, y_d)\pi_1(y_1) \cdots \pi_d(y_d) \neq 0$.

L'estimateur IS est $X_{\text{is}} = h(Y_1, \ldots, Y_d) L(Y_1, \ldots, Y_d)$ où

$$L(y_1, \ldots, y_d) = \frac{\pi_1(y_1) \cdots \pi_d(y_d)}{g_1(y_1) \cdots g_d(y_d)}$$

est le rapport de vraisemblance associé au changement de densité.

Si les Y_j sont dépendants?

Si les Y_j sont dépendants? Et si leur nombre est aléatoire?

Si les Y_j sont dépendants? Et si leur nombre est aléatoire?

Soit $\mu = \mathbb{E}[X]$ où $X = h(Y_1, \ldots, Y_T)$,
$h(Y_1, \ldots, Y_T) = 0$ si $T = \infty$,
et T est un temps d'arrêt par rapport à $\{Y_j,\, j \geq 1\}$.

Si les Y_j sont dépendants? Et si leur nombre est aléatoire?

Soit $\mu = \mathbb{E}[X]$ où $X = h(Y_1, \ldots, Y_T)$,
$h(Y_1, \ldots, Y_T) = 0$ si $T = \infty$,
et T est un temps d'arrêt par rapport à $\{Y_j,\, j \geq 1\}$.

Supposons que la densité de Y_1 est π_1, et la densité Y_j conditionnelle à $(Y_1, \ldots, Y_{j-1}) = (y_1, \ldots, y_{j-1})$ est $\pi_j(\cdot \mid y_1, \ldots, y_{j-1})$.

Si les Y_j sont dépendants? Et si leur nombre est aléatoire?

Soit $\mu = \mathbb{E}[X]$ où $X = h(Y_1, \ldots, Y_T)$,
$h(Y_1, \ldots, Y_T) = 0$ si $T = \infty$,
et T est un temps d'arrêt par rapport à $\{Y_j,\, j \geq 1\}$.

Supposons que la densité de Y_1 est π_1, et la densité Y_j conditionnelle à $(Y_1, \ldots, Y_{j-1}) = (y_1, \ldots, y_{j-1})$ est $\pi_j(\cdot \mid y_1, \ldots, y_{j-1})$.

On remplace les π_j par des densités conditionnelles g_j et on obtient:

$$X_{\mathrm{is}} = h(Y_1, \ldots, Y_T) L(Y_1, \ldots, Y_T)$$

où

$$L(y_1, \ldots, y_T) = \frac{\pi_1(y_1)\pi_2(y_2 \mid y_1) \cdots \pi_T(y_T \mid y_1, \ldots, y_{T-1})}{g_1(y_1)g_2(y_2 \mid y_1) \cdots g_T(y_T \mid y_1, \ldots, y_{T-1})}$$

lorsque $T < \infty$.

Si les Y_j sont dépendants? Et si leur nombre est aléatoire?

Soit $\mu = \mathbb{E}[X]$ où $X = h(Y_1, \ldots, Y_T)$,
$h(Y_1, \ldots, Y_T) = 0$ si $T = \infty$,
et T est un temps d'arrêt par rapport à $\{Y_j,\ j \geq 1\}$.

Supposons que la densité de Y_1 est π_1, et la densité Y_j conditionnelle à $(Y_1, \ldots, Y_{j-1}) = (y_1, \ldots, y_{j-1})$ est $\pi_j(\cdot \mid y_1, \ldots, y_{j-1})$.

On remplace les π_j par des densités conditionnelles g_j et on obtient:

$$X_{\text{is}} = h(Y_1, \ldots, Y_T) L(Y_1, \ldots, Y_T)$$

où

$$L(y_1, \ldots, y_T) = \frac{\pi_1(y_1)\pi_2(y_2 \mid y_1) \cdots \pi_T(y_T \mid y_1, \ldots, y_{T-1})}{g_1(y_1)g_2(y_2 \mid y_1) \cdots g_T(y_T \mid y_1, \ldots, y_{T-1})}$$

lorsque $T < \infty$.

Ici encore, les g_j doivent être tels que L doit être fini lorsque $h(y_1, \ldots, y_T)\pi_1(y_1)\pi_2(y_2 \mid y_1) \cdots \pi_T(y_T \mid y_1, \ldots, y_{T-1}) \neq 0$.

On vérifie que X_{is} est sans biais pour μ:

$$\mu = \mathbb{E}_\pi[h(Y_1, \ldots, Y_T)]$$

On vérifie que X_{is} est sans biais pour μ:

$$\begin{aligned}\mu &= \mathbb{E}_\pi[h(Y_1,\ldots,Y_T)] \\ &= \mathbb{E}_\pi\left[\sum_{n=1}^{\infty} \mathbb{I}[T=n]\, h(Y_1,\ldots,Y_n)\right]\end{aligned}$$

On vérifie que X_{is} est sans biais pour μ:

$$
\begin{aligned}
\mu &= \mathbb{E}_\pi[h(Y_1, \ldots, Y_T)] \\
&= \mathbb{E}_\pi\left[\sum_{n=1}^{\infty} \mathbb{I}[T = n]\, h(Y_1, \ldots, Y_n)\right] \\
&= \sum_{n=1}^{\infty} \mathbb{E}_\pi[\mathbb{I}[T = n]\, h(Y_1, \ldots, Y_n)] \qquad \text{(e.g., si } h \geq 0\text{)}
\end{aligned}
$$

On vérifie que $X_{\rm is}$ est sans biais pour μ:

$$
\begin{aligned}
\mu &= \mathbb{E}_\pi[h(Y_1, \ldots, Y_T)] \\
&= \mathbb{E}_\pi\left[\sum_{n=1}^{\infty} \mathbb{I}[T = n]\, h(Y_1, \ldots, Y_n)\right] \\
&= \sum_{n=1}^{\infty} \mathbb{E}_\pi[\mathbb{I}[T = n]\, h(Y_1, \ldots, Y_n)] \qquad \text{(e.g., si } h \geq 0\text{)} \\
&= \sum_{n=1}^{\infty} \mathbb{E}_g[\mathbb{I}[T = n]\, h(Y_1, \ldots, Y_n) L(Y_1, \ldots, Y_n)]
\end{aligned}
$$

On vérifie que $X_{\rm is}$ est sans biais pour μ:

$$\begin{aligned}
\mu &= \mathbb{E}_\pi[h(Y_1,\ldots,Y_T)] \\
&= \mathbb{E}_\pi\left[\sum_{n=1}^{\infty} \mathbb{I}[T=n]\, h(Y_1,\ldots,Y_n)\right] \\
&= \sum_{n=1}^{\infty} \mathbb{E}_\pi[\mathbb{I}[T=n]\, h(Y_1,\ldots,Y_n)] \qquad \text{(e.g., si } h \ge 0) \\
&= \sum_{n=1}^{\infty} \mathbb{E}_g[\mathbb{I}[T=n]\, h(Y_1,\ldots,Y_n) L(Y_1,\ldots,Y_n)] \\
&= \mathbb{E}_g[h(Y_1,\ldots,Y_T) L(Y_1,\ldots,Y_T)],
\end{aligned}$$

car T est un temps d'arrêt.

On vérifie que $X_{\rm is}$ est sans biais pour μ:

$$\begin{aligned}
\mu &= \mathbb{E}_\pi[h(Y_1,\ldots,Y_T)] \\
&= \mathbb{E}_\pi\left[\sum_{n=1}^{\infty} \mathbb{I}[T=n]\, h(Y_1,\ldots,Y_n)\right] \\
&= \sum_{n=1}^{\infty} \mathbb{E}_\pi[\mathbb{I}[T=n]\, h(Y_1,\ldots,Y_n)] \qquad \text{(e.g., si } h \geq 0\text{)} \\
&= \sum_{n=1}^{\infty} \mathbb{E}_g[\mathbb{I}[T=n]\, h(Y_1,\ldots,Y_n) L(Y_1,\ldots,Y_n)] \\
&= \mathbb{E}_g[h(Y_1,\ldots,Y_T) L(Y_1,\ldots,Y_T)],
\end{aligned}$$

car T est un temps d'arrêt.

Idem pour le cas discret.

Formulation générale

Soit

$$\mu = \mathbb{E}_P[X] = \mathbb{E}_P[h(\omega)] = \int_\Omega h(\omega) dP(\omega)$$

où $P[\omega \in A] = P[A] = \int_A dP(\omega)$ pour tout ensemble mesurable $A \subseteq \Omega$.

Formulation générale

Soit

$$\mu = \mathbb{E}_P[X] = \mathbb{E}_P[h(\omega)] = \int_\Omega h(\omega) dP(\omega)$$

où $P[\omega \in A] = P[A] = \int_A dP(\omega)$ pour tout ensemble mesurable $A \subseteq \Omega$.

On remplace P par une mesure Q telle que $Q(A) > 0$ lorsque $\int_A h(\omega) dP(\omega) > 0$.

Formulation générale

Soit

$$\mu = \mathbb{E}_P[X] = \mathbb{E}_P[h(\omega)] = \int_\Omega h(\omega) dP(\omega)$$

où $P[\omega \in A] = P[A] = \int_A dP(\omega)$ pour tout ensemble mesurable $A \subseteq \Omega$.

On remplace P par une mesure Q telle que $Q(A) > 0$ lorsque $\int_A h(\omega) dP(\omega) > 0$.

Le rapport de vraisemblance est $L(P, Q, \omega) = (dP/dQ)(\omega)$
(c'est la dérivée de Radon-Nikodym de P par rapport à Q) et

$$\mu = \mathbb{E}_P[h(\omega)]$$

Formulation générale

Soit

$$\mu = \mathbb{E}_P[X] = \mathbb{E}_P[h(\omega)] = \int_\Omega h(\omega)dP(\omega)$$

où $P[\omega \in A] = P[A] = \int_A dP(\omega)$ pour tout ensemble mesurable $A \subseteq \Omega$.

On remplace P par une mesure Q telle que $Q(A) > 0$ lorsque $\int_A h(\omega)dP(\omega) > 0$.

Le rapport de vraisemblance est $L(P, Q, \omega) = (dP/dQ)(\omega)$
(c'est la dérivée de Radon-Nikodym de P par rapport à Q) et

$$\mu = \mathbb{E}_P[h(\omega)] = \int_\Omega h(\omega)dP(\omega)$$

Formulation générale

Soit

$$\mu = \mathbb{E}_P[X] = \mathbb{E}_P[h(\omega)] = \int_\Omega h(\omega) dP(\omega)$$

où $P[\omega \in A] = P[A] = \int_A dP(\omega)$ pour tout ensemble mesurable $A \subseteq \Omega$.

On remplace P par une mesure Q telle que $Q(A) > 0$ lorsque $\int_A h(\omega) dP(\omega) > 0$.

Le rapport de vraisemblance est $L(P, Q, \omega) = (dP/dQ)(\omega)$
(c'est la dérivée de Radon-Nikodym de P par rapport à Q) et

$$\mu = \mathbb{E}_P[h(\omega)] = \int_\Omega h(\omega) dP(\omega) = \int_\Omega [h(\omega)(dP/dQ)(\omega)] \, dQ(\omega)$$

Formulation générale

Soit

$$\mu = \mathbb{E}_P[X] = \mathbb{E}_P[h(\omega)] = \int_\Omega h(\omega) dP(\omega)$$

où $P[\omega \in A] = P[A] = \int_A dP(\omega)$ pour tout ensemble mesurable $A \subseteq \Omega$.

On remplace P par une mesure Q telle que $Q(A) > 0$ lorsque $\int_A h(\omega) dP(\omega) > 0$.

Le rapport de vraisemblance est $L(P, Q, \omega) = (dP/dQ)(\omega)$
(c'est la dérivée de Radon-Nikodym de P par rapport à Q) et

$$\begin{aligned} \mu &= \mathbb{E}_P[h(\omega)] = \int_\Omega h(\omega) dP(\omega) = \int_\Omega [h(\omega)(dP/dQ)(\omega)]\, dQ(\omega) \\ &= \mathbb{E}_Q\left[h(\omega) L(P, Q, \omega)\right]. \end{aligned}$$

Formulation générale

Soit

$$\mu = \mathbb{E}_P[X] = \mathbb{E}_P[h(\omega)] = \int_\Omega h(\omega) dP(\omega)$$

où $P[\omega \in A] = P[A] = \int_A dP(\omega)$ pour tout ensemble mesurable $A \subseteq \Omega$.

On remplace P par une mesure Q telle que $Q(A) > 0$ lorsque $\int_A h(\omega) dP(\omega) > 0$.

Le rapport de vraisemblance est $L(P, Q, \omega) = (dP/dQ)(\omega)$
(c'est la dérivée de Radon-Nikodym de P par rapport à Q) et

$$\begin{aligned}
\mu &= \mathbb{E}_P[h(\omega)] = \int_\Omega h(\omega) dP(\omega) = \int_\Omega [h(\omega)(dP/dQ)(\omega)]\, dQ(\omega) \\
&= \mathbb{E}_Q\left[h(\omega) L(P, Q, \omega)\right].
\end{aligned}$$

L'estimateur IS $X_{\mathrm{is}} = h(\omega) L(P, Q, \omega)$ est sans biais pour μ.

Le Q optimal est $Q^*(d\omega) = |h(\omega)|P(d\omega)/\tilde{\mu}$, où $\tilde{\mu} = \int_\Omega |h(\omega)|dP(\omega)$.

Le Q optimal est $Q^*(d\omega) = |h(\omega)|P(d\omega)/\tilde{\mu}$, où $\tilde{\mu} = \int_\Omega |h(\omega)| dP(\omega)$.

Avec ce Q^*, on obtient $X^*_{\mathrm{is}} = \tilde{\mu}$ si $h(\omega) > 0$, $X^*_{\mathrm{is}} = -\tilde{\mu}$ si $h(\omega) < 0$, et $Q^*[h(\omega) = 0] = 0$.

Le Q optimal est $Q^*(d\omega) = |h(\omega)|P(d\omega)/\tilde{\mu}$, où $\tilde{\mu} = \int_\Omega |h(\omega)|dP(\omega)$.

Avec ce Q^*, on obtient $X^*_{\rm is} = \tilde{\mu}$ si $h(\omega) > 0$, $X^*_{\rm is} = -\tilde{\mu}$ si $h(\omega) < 0$, et $Q^*[h(\omega) = 0] = 0$.

Si $h(\omega)$ est toujours du même signe, Q^* réduit la variance à zéro. Mais Q^* est habituellement trop difficile à trouver et à utiliser en pratique.

Le rapport de vraisemblance n'est jamais négatif. De plus,

Proposition.
Si $(dP/dQ)(\omega)$ existe sur un ensemble mesurable B tel que $\mathbb{Q}[B] = 1$, alors $\mathbb{E}_{\mathbb{Q}}[L(\mathbb{P}, \mathbb{Q}, \omega)] = \mathbb{P}[B]$.

Preuve:

$$\mathbb{E}_{\mathbb{Q}}[L(\mathbb{P}, \mathbb{Q}, \omega)] = \int_B [(d\mathbb{P}/d\mathbb{Q})(\omega)] d\mathbb{Q}(\omega) = \int_B d\mathbb{P}(\omega) = \mathbb{P}[B].$$

La décomposition suivante de $\mathrm{Var}[X]$ indique que le rapport de vraisemblance $L(P, Q, \omega)$ doit être petit lorsque $h^2(\omega)$ est grand. Le terme de droite représente la réduction de variance.

La décomposition suivante de $\mathrm{Var}[X]$ indique que le rapport de vraisemblance $L(P, Q, \omega)$ doit être petit lorsque $h^2(\omega)$ est grand. Le terme de droite représente la réduction de variance.

Proposition. $\mathrm{Var}[X] = \mathrm{Var}[X_{\mathrm{is}}] + \mathbb{E}_P[h^2(\omega)(1 - L(P, Q, \omega))]$.

La décomposition suivante de $\mathrm{Var}[X]$ indique que le rapport de vraisemblance $L(P,Q,\omega)$ doit être petit lorsque $h^2(\omega)$ est grand. Le terme de droite représente la réduction de variance.

Proposition. $\mathrm{Var}[X] = \mathrm{Var}[X_{\mathrm{is}}] + \mathbb{E}_P[h^2(\omega)(1 - L(P,Q,\omega))]$.

Preuve.

$$\mathbb{E}_Q[X_{\mathrm{is}}^2] \quad = \quad \int_\Omega [h(\omega)L(P,Q,\omega)]^2 dQ(\omega)$$

La décomposition suivante de $\mathrm{Var}[X]$ indique que le rapport de vraisemblance $L(P,Q,\omega)$ doit être petit lorsque $h^2(\omega)$ est grand. Le terme de droite représente la réduction de variance.

Proposition. $\mathrm{Var}[X] = \mathrm{Var}[X_{\mathrm{is}}] + \mathbb{E}_P[h^2(\omega)(1 - L(P,Q,\omega))]$.

Preuve.

$$\begin{aligned} \mathbb{E}_Q[X_{\mathrm{is}}^2] &= \int_\Omega [h(\omega)L(P,Q,\omega)]^2 dQ(\omega) \\ &= \int_\Omega h^2(\omega)L(P,Q,\omega)dP(\omega) \end{aligned}$$

La décomposition suivante de $\mathrm{Var}[X]$ indique que le rapport de vraisemblance $L(P,Q,\omega)$ doit être petit lorsque $h^2(\omega)$ est grand. Le terme de droite représente la réduction de variance.

Proposition. $\mathrm{Var}[X] = \mathrm{Var}[X_{\mathrm{is}}] + \mathbb{E}_P[h^2(\omega)(1 - L(P,Q,\omega))]$.

Preuve.

$$\begin{aligned}
\mathbb{E}_Q[X_{\mathrm{is}}^2] &= \int_\Omega [h(\omega)L(P,Q,\omega)]^2 dQ(\omega) \\
&= \int_\Omega h^2(\omega)L(P,Q,\omega)dP(\omega) \\
&= \mathbb{E}_P[h^2(\omega)L(P,Q,\omega)].
\end{aligned}$$

La décomposition suivante de $\mathrm{Var}[X]$ indique que le rapport de vraisemblance $L(P, Q, \omega)$ doit être petit lorsque $h^2(\omega)$ est grand. Le terme de droite représente la réduction de variance.

Proposition. $\mathrm{Var}[X] = \mathrm{Var}[X_{\mathrm{is}}] + \mathbb{E}_P[h^2(\omega)(1 - L(P, Q, \omega))]$.

Preuve.

$$\begin{aligned} \mathbb{E}_Q[X_{\mathrm{is}}^2] &= \int_\Omega [h(\omega)L(P, Q, \omega)]^2 dQ(\omega) \\ &= \int_\Omega h^2(\omega)L(P, Q, \omega)dP(\omega) \\ &= \mathbb{E}_P[h^2(\omega)L(P, Q, \omega)]. \end{aligned}$$

On a donc

$$\mathrm{Var}[X] - \mathrm{Var}[X_{\mathrm{is}}]$$

La décomposition suivante de $\mathrm{Var}[X]$ indique que le rapport de vraisemblance $L(P, Q, \omega)$ doit être petit lorsque $h^2(\omega)$ est grand. Le terme de droite représente la réduction de variance.

Proposition. $\mathrm{Var}[X] = \mathrm{Var}[X_{\mathrm{is}}] + \mathbb{E}_P[h^2(\omega)(1 - L(P, Q, \omega))]$.

Preuve.

$$\begin{aligned}
\mathbb{E}_Q[X_{\mathrm{is}}^2] &= \int_\Omega [h(\omega)L(P, Q, \omega)]^2 dQ(\omega) \\
&= \int_\Omega h^2(\omega)L(P, Q, \omega)dP(\omega) \\
&= \mathbb{E}_P[h^2(\omega)L(P, Q, \omega)].
\end{aligned}$$

On a donc

$$\mathrm{Var}[X] - \mathrm{Var}[X_{\mathrm{is}}] = \mathbb{E}_P[X^2] - \mathbb{E}_Q[X_{\mathrm{is}}^2]$$

La décomposition suivante de $\mathrm{Var}[X]$ indique que le rapport de vraisemblance $L(P,Q,\omega)$ doit être petit lorsque $h^2(\omega)$ est grand. Le terme de droite représente la réduction de variance.

Proposition. $\mathrm{Var}[X] = \mathrm{Var}[X_{\mathrm{is}}] + \mathbb{E}_P[h^2(\omega)(1 - L(P,Q,\omega))]$.

Preuve.

$$\begin{aligned}\mathbb{E}_Q[X_{\mathrm{is}}^2] &= \int_\Omega [h(\omega)L(P,Q,\omega)]^2 dQ(\omega)\\ &= \int_\Omega h^2(\omega)L(P,Q,\omega)dP(\omega)\\ &= \mathbb{E}_P[h^2(\omega)L(P,Q,\omega)].\end{aligned}$$

On a donc

$$\mathrm{Var}[X] - \mathrm{Var}[X_{\mathrm{is}}] = \mathbb{E}_P[X^2] - \mathbb{E}_Q[X_{\mathrm{is}}^2] = \mathbb{E}_P[h^2(\omega)(1 - L(P,Q,\omega))].$$

Condition suffisante de réduction de variance par un facteur $\rho \leq 1$.

Condition suffisante de réduction de variance par un facteur $\rho \leq 1$.

Corollaire. Si $L(P, Q, \omega) \leq \rho$ lorsque $h(\omega) \neq 0$, pour une constante $\rho \leq 1$, alors

$$\mathrm{Var}[X_{\mathrm{is}}] \leq \rho \mathrm{Var}[X] - (1 - \rho)\mu^2$$

Condition suffisante de réduction de variance par un facteur $\rho \leq 1$.

Corollaire. Si $L(P, Q, \omega) \leq \rho$ lorsque $h(\omega) \neq 0$, pour une constante $\rho \leq 1$, alors

$$\mathrm{Var}[X_{\mathrm{is}}] \leq \rho \mathrm{Var}[X] - (1-\rho)\mu^2 \leq \rho \mathrm{Var}[X].$$

Condition suffisante de réduction de variance par un facteur $\rho \leq 1$.

Corollaire. Si $L(P, Q, \omega) \leq \rho$ lorsque $h(\omega) \neq 0$, pour une constante $\rho \leq 1$, alors

$$\mathrm{Var}[X_{\mathrm{is}}] \leq \rho \mathrm{Var}[X] - (1-\rho)\mu^2 \leq \rho \mathrm{Var}[X].$$

Preuve

$$\begin{aligned} \mathrm{Var}[X] - \mathrm{Var}[X_{\mathrm{is}}] &= \mathbb{E}_P[h^2(\omega)(1 - L(\omega))] \\ &\geq (1-\rho)\mathbb{E}_P[h^2(\omega)] \\ &\geq (1-\rho)(\mathrm{Var}[X] + \mu^2) \end{aligned}$$

Condition suffisante de réduction de variance par un facteur $\rho \le 1$.

Corollaire. Si $L(P, Q, \omega) \le \rho$ lorsque $h(\omega) \neq 0$, pour une constante $\rho \le 1$, alors

$$\mathrm{Var}[X_{\mathrm{is}}] \le \rho \mathrm{Var}[X] - (1-\rho)\mu^2 \le \rho \mathrm{Var}[X].$$

Preuve

$$\begin{aligned}
\mathrm{Var}[X] - \mathrm{Var}[X_{\mathrm{is}}] &= \mathbb{E}_P[h^2(\omega)(1 - L(\omega))] \\
&\ge (1-\rho)\mathbb{E}_P[h^2(\omega)] \\
&\ge (1-\rho)(\mathrm{Var}[X] + \mu^2)
\end{aligned}$$

Interprétation: On veut choisir Q de manière à ce que les grandes valeurs de $L(P, Q, \omega)$ surviennent lorsque $h(\omega) = 0$ (ou proche).
Ainsi, L sera plus petit (i.e., ω plus “probable” sous Q) lorsque $h^2(\omega)$ est grand.

Condition suffisante de réduction de variance par un facteur $\rho \le 1$.

Corollaire. Si $L(P, Q, \omega) \le \rho$ lorsque $h(\omega) \ne 0$, pour une constante $\rho \le 1$, alors

$$\mathrm{Var}[X_{\mathrm{is}}] \le \rho \mathrm{Var}[X] - (1-\rho)\mu^2 \le \rho \mathrm{Var}[X].$$

Preuve

$$\begin{aligned} \mathrm{Var}[X] - \mathrm{Var}[X_{\mathrm{is}}] &= \mathbb{E}_P[h^2(\omega)(1 - L(\omega))] \\ &\ge (1-\rho)\mathbb{E}_P[h^2(\omega)] \\ &\ge (1-\rho)(\mathrm{Var}[X] + \mu^2) \end{aligned}$$

Interprétation: On veut choisir Q de manière à ce que les grandes valeurs de $L(P, Q, \omega)$ surviennent lorsque $h(\omega) = 0$ (ou proche).
Ainsi, L sera plus petit (i.e., ω plus "probable" sous Q) lorsque $h^2(\omega)$ est grand.

Exemple. Si $h(\omega) = \mathbb{I}[\omega \in A]$ et on veut estimer $\mu = P[A] = \mathbb{E}_P[h(\omega)]$, Q^* est défini par $Q^*(B) = \mathbb{I}[A]P[B]/P[A] = P[B \mid A]$ si $B \subseteq A$.
On alors $L(\omega) = (\mathrm{d}P(\omega)/\mathrm{d}Q^*(\omega)) = P[A] = \rho$ pour $\omega \in A$.

Variance zéro pour une chaîne de Markov

CMTC $\{Y_j,\, j \ge 0\}$ dans $\mathcal{Y}$,
noyau de transition $\mathbb{P}$, fonction de coût $c : \mathcal{Y}^2 \to [0, \infty)$.

Dans l'état $Y_{j-1} = y \in \mathcal{Y}$, la loi de prob. du prochain état Y_j est
$\mathbb{P}[Y_j \in \cdot \mid Y_{j-1} = y) = \mathbb{P}(\cdot \mid y)$, et on paye $c(Y_{j-1}, Y_j)$ à l'étape j.

États absorbants: $\Delta \subset \mathcal{Y}$. On a $\mathbb{P}(\{y\} \mid y) = 1$ et $c(y, y) = 0$ pour tout $y \in \Delta$.

Variance zéro pour une chaîne de Markov

CMTC $\{Y_j,\ j \geq 0\}$ dans $\mathcal{Y}$,
noyau de transition $\mathbb{P}$, fonction de coût $c : \mathcal{Y}^2 \to [0, \infty)$.

Dans l'état $Y_{j-1} = y \in \mathcal{Y}$, la loi de prob. du prochain état Y_j est
$\mathbb{P}[Y_j \in \cdot \mid Y_{j-1} = y) = \mathbb{P}(\cdot \mid y)$, et on paye $c(Y_{j-1}, Y_j)$ à l'étape j.

États absorbants: $\Delta \subset \mathcal{Y}$. On a $\mathbb{P}(\{y\} \mid y) = 1$ et $c(y, y) = 0$ pour tout $y \in \Delta$.

Soit $\tau = \inf\{j : Y_j \in \Delta\}$,

$$X = \sum_{j=1}^{\tau} c(Y_{j-1}, Y_j) \quad \text{et} \quad \mu(y) = \mathbb{E}[X \mid Y_0 = y],$$

le coût total espéré en partant de y.
On suppose que $\mathbb{E}[\tau \mid Y_0 = y] < \infty$ et $\mu(y) < \infty$ pour tout $y \in \mathcal{Y}$.

Variance zéro pour une chaîne de Markov

CMTC $\{Y_j,\ j \geq 0\}$ dans $\mathcal{Y}$,
noyau de transition $\mathbb{P}$, fonction de coût $c : \mathcal{Y}^2 \to [0, \infty)$.

Dans l'état $Y_{j-1} = y \in \mathcal{Y}$, la loi de prob. du prochain état Y_j est $\mathbb{P}[Y_j \in \cdot \mid Y_{j-1} = y) = \mathbb{P}(\cdot \mid y)$, et on paye $c(Y_{j-1}, Y_j)$ à l'étape j.

États absorbants: $\Delta \subset \mathcal{Y}$. On a $\mathbb{P}(\{y\} \mid y) = 1$ et $c(y, y) = 0$ pour tout $y \in \Delta$.

Soit $\tau = \inf\{j : Y_j \in \Delta\}$,

$$X = \sum_{j=1}^{\tau} c(Y_{j-1}, Y_j) \quad \text{et} \quad \mu(y) = \mathbb{E}[X \mid Y_0 = y],$$

le coût total espéré en partant de y.
On suppose que $\mathbb{E}[\tau \mid Y_0 = y] < \infty$ et $\mu(y) < \infty$ pour tout $y \in \mathcal{Y}$.

La fonction $\mu : \mathcal{Y} \to [0, \infty)$ satisfait la récurrence

$$\begin{aligned} \mu(y) &= \mathbb{E}[c(y, Y_1) + \mu(Y_1) \mid Y_0 = y] \\ &= \int_{\mathcal{Y}} [c(y, y_1) + \mu(y_1)] d\mathbb{P}(y_1 \mid y). \end{aligned}$$

On change $\mathbb{P}$ pour $\mathbb{Q}$ tel que $\mathbb{Q}(B \mid y) > 0$ pour tout B tel que $\int_B [c(y, y_1) + \mu(y_1)] d\mathbb{P}(y_1 \mid y) > 0$. L'estimateur devient:

$$X_{\text{is}} = \sum_{j=1}^{\tau} c(Y_{j-1}, Y_j) \prod_{i=1}^{j} L(Y_{i-1}, Y_i),$$

où $L(Y_{i-1}, Y_i) = (d\mathbb{P}/d\mathbb{Q})(Y_i \mid Y_{i-1})$. On a $\mathbb{E}_{\mathbb{Q},y}[X_{\text{is}}] = \mu(y)$.

On change $\mathbb{P}$ pour $\mathbb{Q}$ tel que $\mathbb{Q}(B \mid y) > 0$ pour tout B tel que $\int_B [c(y, y_1) + \mu(y_1)] d\mathbb{P}(y_1 \mid y) > 0$. L'estimateur devient:

$$X_{\mathrm{is}} = \sum_{j=1}^{\tau} c(Y_{j-1}, Y_j) \prod_{i=1}^{j} L(Y_{i-1}, Y_i),$$

où $L(Y_{i-1}, Y_i) = (d\mathbb{P}/d\mathbb{Q})(Y_i \mid Y_{i-1})$. On a $\mathbb{E}_{\mathbb{Q},y}[X_{\mathrm{is}}] = \mu(y)$.

On définit

$$\mathbb{Q}^*(dy_1 \mid y) = \mathbb{P}(dy_1 \mid y) \frac{c(y, y_1) + \mu(y_1)}{\mu(y)} \quad \text{si } \mu(y) > 0$$

et $\mathbb{Q}^*(\cdot \mid y) = \mathbb{P}(\cdot \mid y)$ si $\mu(y) = 0$.

On change $\mathbb{P}$ pour $\mathbb{Q}$ tel que $\mathbb{Q}(B \mid y) > 0$ pour tout B tel que $\int_B [c(y, y_1) + \mu(y_1)] d\mathbb{P}(y_1 \mid y) > 0$. L'estimateur devient:

$$X_{\text{is}} = \sum_{j=1}^{\tau} c(Y_{j-1}, Y_j) \prod_{i=1}^{j} L(Y_{i-1}, Y_i),$$

où $L(Y_{i-1}, Y_i) = (d\mathbb{P}/d\mathbb{Q})(Y_i \mid Y_{i-1})$. On a $\mathbb{E}_{\mathbb{Q},y}[X_{\text{is}}] = \mu(y)$.

On définit

$$\mathbb{Q}^*(dy_1 \mid y) = \mathbb{P}(dy_1 \mid y) \frac{c(y, y_1) + \mu(y_1)}{\mu(y)} \quad \text{si } \mu(y) > 0$$

et $\mathbb{Q}^*(\cdot \mid y) = \mathbb{P}(\cdot \mid y)$ si $\mu(y) = 0$.

Proposition. $\mathbb{Q} = \mathbb{Q}^*$ donne une variance de zéro: $\mathrm{Var}_{\mathbb{Q}^*,y}[X_{\text{is}}] = 0$.

Habituellement pas implantable directement, mais peut donner une idée de quoi faire si on dispose d'une approximation de μ.

Exemple: option asiatique sous MBG

On veut estimer $v(s_0, T) = \mathbb{E}[Y(s_0)]$, où

$$\begin{aligned} Y(s_0) &= e^{-rT} \max\left(0,\ s_0 W - K\right), \\ W &= \frac{1}{d} \sum_{i=1}^{d} \exp\left[(r - \sigma^2/2) t_i + \sigma \sum_{j=1}^{i} \sqrt{t_j - t_{j-1}} Z_j \right] \end{aligned}$$

et $Z_1, \ldots, Z_d$ sont i.i.d. $N(0,1)$.

Exemple: option asiatique sous MBG

On veut estimer $v(s_0, T) = \mathbb{E}[Y(s_0)]$, où

$$\begin{aligned} Y(s_0) &= e^{-rT} \max\left(0,\ s_0 W - K\right), \\ W &= \frac{1}{d} \sum_{i=1}^{d} \exp\left[(r - \sigma^2/2) t_i + \sigma \sum_{j=1}^{i} \sqrt{t_j - t_{j-1}} Z_j \right] \end{aligned}$$

et $Z_1, \dots, Z_d$ sont $\text{i.i.d. } N(0,1)$.

Si $K \gg s_0 \exp[r - \sigma^2/2]$, on veut augmenter le “drift”.
Idée: augmenter la moyenne de Z_j de 0 à ν_j. Ceci ajoute $\delta_i = \sigma \sum_{j=1}^{i} \nu_j \sqrt{t_j - t_{j-1}}$ à l'exposant de l'exponentielle. On a

$$L(\omega) = \prod_{j=1}^{d} \frac{\exp\left(-Z_j^2/2\right)}{\exp\left(-(Z_j - \nu_j)^2/2\right)} = \exp\left(\sum_{j=1}^{d} (\nu_j^2/2 - \nu_j Z_j) \right).$$

Implantation: générer les $Z_j \sim \text{Normale}(0,1)$, ajouter δ_i à l'exposant pour chaque i, puis multiplier $Y(s_0)$ par $L(\omega)$. Habituellement, on prend les ν_j égaux.

Mais comment choisir ces ν_j?

Exemple.

Une chaine de Markov $\{X_n,\, n \geq 0\}$ sur les états $\{0, 1, \ldots, K\}$, avec $X_0 = 0$.

Probabilités de transition $p_{i,j} = P[X_n = j \mid X_{n-1} = i]$.

On veut estimer la probabilité $\mu = \mu(x_0)$ d'atteindre K avant de revenir à 0.

Exemple.

Une chaine de Markov $\{X_n,\, n \ge 0\}$ sur les états $\{0, 1, \ldots, K\}$, avec $X_0 = 0$.

Probabilités de transition $p_{i,j} = P[X_n = j \mid X_{n-1} = i]$.

On veut estimer la probabilité $\mu = \mu(x_0)$ d'atteindre K avant de revenir à 0.

Estimateur naif: $\mathbb{I}[X_T = K]$ où $T = \inf\{n \ge 1 : X_n \in \{0, K\}\}$.

Exemple.

Une chaine de Markov $\{X_n,\, n \geq 0\}$ sur les états $\{0, 1, \ldots, K\}$, avec $X_0 = 0$.

Probabilités de transition $p_{i,j} = P[X_n = j \mid X_{n-1} = i]$.

On veut estimer la probabilité $\mu = \mu(x_0)$ d'atteindre K avant de revenir à 0.

Estimateur naif: $\mathbb{I}[X_T = K]$ où $T = \inf\{n \geq 1 : X_n \in \{0, K\}\}$.

IS: changer les $p_{i,j}$ pour des $q_{i,j}$ pour augmenter les chances d'aller à K.

Exemple.

Une chaine de Markov $\{X_n,\, n \geq 0\}$ sur les états $\{0, 1, \ldots, K\}$, avec $X_0 = 0$.

Probabilités de transition $p_{i,j} = P[X_n = j \mid X_{n-1} = i]$.

On veut estimer la probabilité $\mu = \mu(x_0)$ d'atteindre K avant de revenir à 0.

Estimateur naif: $\mathbb{I}[X_T = K]$ où $T = \inf\{n \geq 1 : X_n \in \{0, K\}\}$.

IS: changer les $p_{i,j}$ pour des $q_{i,j}$ pour augmenter les chances d'aller à K.

Idée simpliste: bloquer les retours à 0 en posant $q_{i,0} = 0$ pour tout $i > 0$, et renormaliser les autres probabilités: $q_{i,j} = p_{i,j}/(1 - p_{i,0})$ pour $i, j > 0$, et $q_{0,j} = p_{0,j}$ pour tout j.

Exemple.

Une chaine de Markov $\{X_n,\ n \geq 0\}$ sur les états $\{0, 1, \ldots, K\}$, avec $X_0 = 0$.

Probabilités de transition $p_{i,j} = P[X_n = j \mid X_{n-1} = i]$.

On veut estimer la probabilité $\mu = \mu(x_0)$ d'atteindre K avant de revenir à 0.

Estimateur naif: $\mathbb{I}[X_T = K]$ où $T = \inf\{n \geq 1 : X_n \in \{0, K\}\}$.

IS: changer les $p_{i,j}$ pour des $q_{i,j}$ pour augmenter les chances d'aller à K.

Idée simpliste: bloquer les retours à 0 en posant $q_{i,0} = 0$ pour tout $i > 0$, et renormaliser les autres probabilités: $q_{i,j} = p_{i,j}/(1 - p_{i,0})$ pour $i, j > 0$, et $q_{0,j} = p_{0,j}$ pour tout j. On a alors $P[X_T = K] = 1$ et

$$L(\omega) = L(X_1, \ldots, X_T) = \prod_{n=1}^{T} \frac{p_{X_{n-1}, X_n}}{q_{X_{n-1}, X_n}} = \prod_{n=2}^{T} (1 - p_{X_{n-1}, 0}) \leq 1.$$

Donc la variance est réduite.

Mais: chaque simulation risque d'être très longue (on peut de se promener autour de l'état 0 très longtemps). L'efficacité n'est pas nécessairement améliorée.

Exemple.

Une chaine de Markov $\{X_n,\ n \geq 0\}$ sur les états $\{0, 1, \ldots, K\}$, avec $X_0 = 0$.

Probabilités de transition $p_{i,j} = P[X_n = j \mid X_{n-1} = i]$.

On veut estimer la probabilité $\mu = \mu(x_0)$ d'atteindre K avant de revenir à 0.

Estimateur naif: $\mathbb{I}[X_T = K]$ où $T = \inf\{n \geq 1 : X_n \in \{0, K\}\}$.

IS: changer les $p_{i,j}$ pour des $q_{i,j}$ pour augmenter les chances d'aller à K.

Idée simpliste: bloquer les retours à 0 en posant $q_{i,0} = 0$ pour tout $i > 0$, et renormaliser les autres probabilités: $q_{i,j} = p_{i,j}/(1 - p_{i,0})$ pour $i, j > 0$, et $q_{0,j} = p_{0,j}$ pour tout j. On a alors $P[X_T = K] = 1$ et

$$L(\omega) = L(X_1, \ldots, X_T) = \prod_{n=1}^{T} \frac{p_{X_{n-1},X_n}}{q_{X_{n-1},X_n}} = \prod_{n=2}^{T} (1 - p_{X_{n-1},0}) \leq 1.$$

Donc la variance est réduite.

Mais: chaque simulation risque d'être très longue (on peut de se promener autour de l'état 0 très longtemps). L'efficacité n'est pas nécessairement améliorée.
Variance zéro pour cet exemple?

Une marche aléatoire unidimensionnelle sur $\{0, 1, \ldots, K\}$.

Soient $p_{i,i+1} = p$ et $p_{i,i-1} = 1 - p$ pour $1 \leq i \leq K - 1$, et $p_{0,1} = p_{K,K-1} = 1$.

Une marche aléatoire unidimensionnelle sur $\{0, 1, \ldots, K\}$.

Soient $p_{i,i+1} = p$ et $p_{i,i-1} = 1 - p$ pour $1 \le i \le K - 1$, et $p_{0,1} = p_{K,K-1} = 1$.

Si $p < 1/2$ et K est grand, la chaine est attirée vers 0.

Couper l'accès à 0 ne suffit pas, il faut augmenter l'attirance vers K.

Une marche aléatoire unidimensionnelle sur $\{0, 1, \ldots, K\}$.

Soient $p_{i,i+1} = p$ et $p_{i,i-1} = 1 - p$ pour $1 \leq i \leq K - 1$, et $p_{0,1} = p_{K,K-1} = 1$.

Si $p < 1/2$ et K est grand, la chaine est attirée vers 0.

Couper l'accès à 0 ne suffit pas, il faut augmenter l'attirance vers K.

On va modifier les probabilités pour $q_{i,i+1} = q$ et $q_{i,i-1} = 1 - q$, pour $q > p$.

Une marche aléatoire unidimensionnelle sur $\{0, 1, \ldots, K\}$.

Soient $p_{i,i+1} = p$ et $p_{i,i-1} = 1 - p$ pour $1 \leq i \leq K - 1$, et $p_{0,1} = p_{K,K-1} = 1$.

Si $p < 1/2$ et K est grand, la chaine est attirée vers 0.

Couper l'accès à 0 ne suffit pas, il faut augmenter l'attirance vers K.

On va modifier les probabilités pour $q_{i,i+1} = q$ et $q_{i,i-1} = 1 - q$, pour $q > p$.

Examinons le rapport de vraisemblance lorsque $X_T = K$.
Pour chaque paire d'états $(i, i+1)$ on va de i à $i+1$ une fois de plus que de $i+1$ à i. On a alors

$$L(\omega) = \prod_{n=1}^{T} \frac{p_{X_{n-1},X_n}}{q_{X_{n-1},X_n}} = \left(\frac{p}{q}\right)^{K-1} \left(\frac{p(1-p)}{q(1-q)}\right)^{(T-K)/2}.$$

$$L(\omega) \;=\; \left(\frac{p}{q}\right)^{K-1} \left(\frac{p(1-p)}{q(1-q)}\right)^{(T-K)/2} .$$

Pour s'assurer que $L(\omega) < 1$, prenons $q > p$ et $q(1-q) \geq p(1-p)$, i.e., $p < q \leq 1 - p$.

$$L(\omega) \;=\; \left(\frac{p}{q}\right)^{K-1} \left(\frac{p(1-p)}{q(1-q)}\right)^{(T-K)/2}.$$

Pour s'assurer que $L(\omega) < 1$, prenons $q > p$ et $q(1-q) \geq p(1-p)$, i.e., $p < q \leq 1-p$.

En maximisant q sous cette contrainte (pour aller à K le plus vite et le plus souvent possible), on obtient $q = 1-p$.

Le RV se simplifie alors et devient une constante:

$$L(\omega) = \left(\frac{p}{q}\right)^{K-1} = \left(\frac{p}{1-p}\right)^{K-1}.$$

Avec IS, la variance de l'estimateur est multipliée par $L(\omega)$.

$$L(\omega) \;=\; \left(\frac{p}{q}\right)^{K-1} \left(\frac{p(1-p)}{q(1-q)}\right)^{(T-K)/2}.$$

Pour s'assurer que $L(\omega) < 1$, prenons $q > p$ et $q(1-q) \geq p(1-p)$, i.e., $p < q \leq 1-p$.

En maximisant q sous cette contrainte (pour aller à K le plus vite et le plus souvent possible), on obtient $q = 1-p$.

Le RV se simplifie alors et devient une constante:

$$L(\omega) = \left(\frac{p}{q}\right)^{K-1} = \left(\frac{p}{1-p}\right)^{K-1}.$$

Avec IS, la variance de l'estimateur est multipliée par $L(\omega)$.

Example: si $p = 1/3$ et $K = 101$, la variance est divisée par $2^{100} \approx 1.2 \times 10^{30}$.

Marche aléatoire sur $\mathbb{R}$

On passe d'une marche sur $\{0, \ldots, K\}$ à une marche sur $\mathbb{R}$.
Ce modèle a de nombreuses applications.

Marche aléatoire sur $\mathbb{R}$

On passe d'une marche sur $\{0, \ldots, K\}$ à une marche sur $\mathbb{R}$.
Ce modèle a de nombreuses applications.

L'état à l'étape n est

$$D_n = \sum_{j=1}^{n} Y_j, \qquad n \geq 0,$$

où les Y_j sont i.i.d. de densité π, avec $\mathbb{E}[Y_j] < 0$.

Marche aléatoire sur $\mathbb{R}$

On passe d'une marche sur $\{0, \ldots, K\}$ à une marche sur $\mathbb{R}$.
Ce modèle a de nombreuses applications.

L'état à l'étape n est

$$D_n = \sum_{j=1}^{n} Y_j, \qquad n \geq 0,$$

où les Y_j sont i.i.d. de densité π, avec $\mathbb{E}[Y_j] < 0$.

Bien sûr, $P[\lim_{n\to\infty} D_n = -\infty] = 1$.

Marche aléatoire sur $\mathbb{R}$

On passe d'une marche sur $\{0, \ldots, K\}$ à une marche sur $\mathbb{R}$.
Ce modèle a de nombreuses applications.

L'état à l'étape n est

$$D_n = \sum_{j=1}^{n} Y_j, \qquad n \geq 0,$$

où les Y_j sont i.i.d. de densité π, avec $\mathbb{E}[Y_j] < 0$.

Bien sûr, $P[\lim_{n\to\infty} D_n = -\infty] = 1$.

D_n part de 0, se promène autour un moment, et s'en va éventuellement vers $-\infty$.

Marche aléatoire sur $\mathbb{R}$

On passe d'une marche sur $\{0, \ldots, K\}$ à une marche sur $\mathbb{R}$.
Ce modèle a de nombreuses applications.

L'état à l'étape n est

$$D_n = \sum_{j=1}^{n} Y_j, \qquad n \geq 0,$$

où les Y_j sont i.i.d. de densité π, avec $\mathbb{E}[Y_j] < 0$.

Bien sûr, $P[\lim_{n\to\infty} D_n = -\infty] = 1$.

D_n part de 0, se promène autour un moment, et s'en va éventuellement vers $-\infty$.

Pour une constante $\ell > 0$, soit

$$T_\ell = \inf\{n \geq 0 : D_n \geq \ell\}.$$

On veut estimer

$$\mu_\ell = P[T_\ell < \infty] = P[\max\{D_n,\, n > 0\} \geq \ell].$$

L'estimateur naif $\mathbb{I}[T_\ell < \infty]$ est peu pratique: si $T_\ell = \infty$, il faut simuler indéfiniment pour en être certain.
Et si ℓ est grand, $\{T_\ell < \infty\}$ est un événement rare.

L'estimateur naif $\mathbb{I}[T_\ell < \infty]$ est peu pratique: si $T_\ell = \infty$, il faut simuler indéfiniment pour en être certain.
Et si ℓ est grand, $\{T_\ell < \infty\}$ est un événement rare.

Le changement de mesure optimal serait la loi conditionnelle à $\{T_\ell < \infty\}$.
Trop compliqué. Mais on va tenter de l'approximer grossièrement.

L'estimateur naif $\mathbb{I}[T_\ell < \infty]$ est peu pratique: si $T_\ell = \infty$, il faut simuler indéfiniment pour en être certain.
Et si ℓ est grand, $\{T_\ell < \infty\}$ est un événement rare.

Le changement de mesure optimal serait la loi conditionnelle à $\{T_\ell < \infty\}$.
Trop compliqué. Mais on va tenter de l'approximer grossièrement.

On va simplement changer la densité π des Y_j par

$$\pi_\theta(y) = \pi(y)\frac{e^{\theta y}}{M(\theta)} = e^{\theta y - \Psi(\theta)}\pi(y), \qquad y \in \mathbb{R},$$

où le paramètre $\theta > 0$ reste à déterminer,

$$M(\theta) = \int_0^\infty e^{\theta y}\pi(y)dy = \mathbb{E}\left[e^{\theta Y_j}\right] > 1$$

est une constante de normalisation et $\Psi(\theta) = \ln M(\theta)$.

Torsion exponentielle ("exponential twisting"): la densité est gonflée par un facteur qui augmente exponentiellement avec y, puis renormalisée.

Torsion exponentielle ("exponential twisting"): la densité est gonflée par un facteur qui augmente exponentiellement avec y, puis renormalisée.

M et Ψ sont la fonction génératrice des moments et la fonction gérératrice des cumulants de Y_j. Attention: n'existe pas toujours (e.g., Pareto, lognormale, etc,). On a

$$\begin{aligned}
\mathbb{E}_\theta[Y_j] &= \Psi'(\theta) = M'(\theta)/M(\theta), \\
\mathrm{Var}_\theta[Y_j] &= \Psi''(\theta), \qquad etc.
\end{aligned}$$

Supposons que $M(\theta) < \infty$ dans un voisinage de $\theta = 0$, i.e., tous les moments de Y_j sont finis.

Torsion exponentielle ("exponential twisting"): la densité est gonflée par un facteur qui augmente exponentiellement avec y, puis renormalisée.

M et Ψ sont la fonction génératrice des moments et la fonction gérératrice des cumulants de Y_j. Attention: n'existe pas toujours (e.g., Pareto, lognormale, etc,). On a

$$\begin{aligned} \mathbb{E}_\theta[Y_j] &= \Psi'(\theta) = M'(\theta)/M(\theta), \\ \mathrm{Var}_\theta[Y_j] &= \Psi''(\theta), \qquad etc. \end{aligned}$$

Supposons que $M(\theta) < \infty$ dans un voisinage de $\theta = 0$, i.e., tous les moments de Y_j sont finis.

Lorsque $\{T_\ell < \infty\}$, le rapport de vraisemblance devient

$$L(\omega) = \prod_{j=1}^{T_\ell} \frac{\pi(Y_j)}{\pi_\theta(Y_j)}$$

Torsion exponentielle ("exponential twisting"): la densité est gonflée par un facteur qui augmente exponentiellement avec y, puis renormalisée.

M et Ψ sont la fonction génératrice des moments et la fonction gérératrice des cumulants de Y_j. Attention: n'existe pas toujours (e.g., Pareto, lognormale, etc,). On a

$$\begin{aligned} \mathbb{E}_\theta[Y_j] &= \Psi'(\theta) = M'(\theta)/M(\theta), \\ \mathrm{Var}_\theta[Y_j] &= \Psi''(\theta), \qquad etc. \end{aligned}$$

Supposons que $M(\theta) < \infty$ dans un voisinage de $\theta = 0$, i.e., tous les moments de Y_j sont finis.

Lorsque $\{T_\ell < \infty\}$, le rapport de vraisemblance devient

$$L(\omega) = \prod_{j=1}^{T_\ell} \frac{\pi(Y_j)}{\pi_\theta(Y_j)} = [M(\theta)]^{T_\ell} \exp\left(-\theta \sum_{j=1}^{T_\ell} Y_j\right)$$

Torsion exponentielle ("exponential twisting"): la densité est gonflée par un facteur qui augmente exponentiellement avec y, puis renormalisée.

M et Ψ sont la fonction génératrice des moments et la fonction gérératrice des cumulants de Y_j. Attention: n'existe pas toujours (e.g., Pareto, lognormale, etc,). On a

$$\begin{aligned} \mathbb{E}_\theta[Y_j] &= \Psi'(\theta) = M'(\theta)/M(\theta), \\ \mathrm{Var}_\theta[Y_j] &= \Psi''(\theta), \qquad etc. \end{aligned}$$

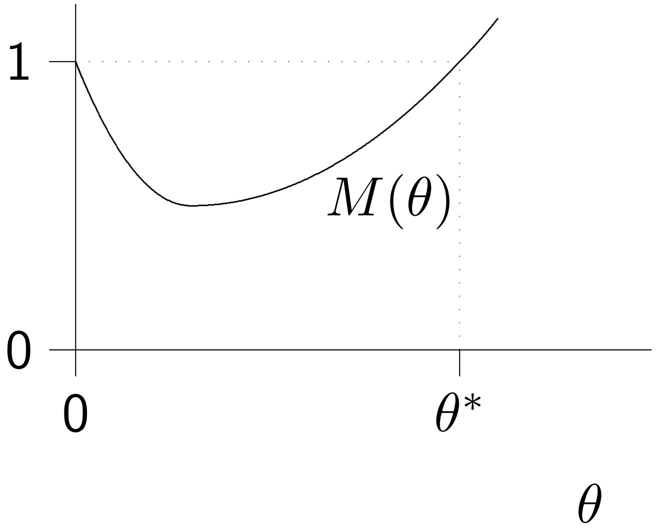
Supposons que $M(\theta) < \infty$ dans un voisinage de $\theta = 0$, i.e., tous les moments de Y_j sont finis.

Lorsque $\{T_\ell < \infty\}$, le rapport de vraisemblance devient

$$L(\omega) = \prod_{j=1}^{T_\ell} \frac{\pi(Y_j)}{\pi_\theta(Y_j)} = [M(\theta)]^{T_\ell} \exp\left(-\theta \sum_{j=1}^{T_\ell} Y_j\right) = [M(\theta)]^{T_\ell} \exp\left(-\theta D_{T_\ell}\right).$$

On sait que $M(\theta)$ est convexe, $M(0) = 1$ et $M'(0) = \mathbb{E}[Y_j] < 0$.

On suppose que $\theta^* \stackrel{\text{def}}{=} \sup\{\theta > 0 : M(\theta) \leq 1\} < \infty$.

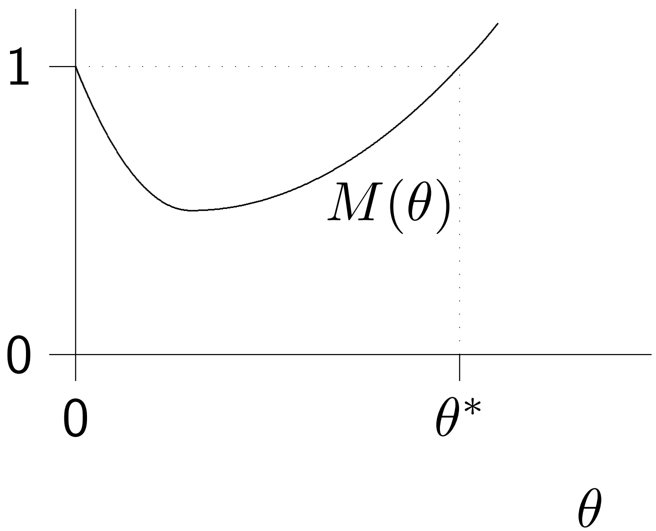


Pour IS, on va prendre $\theta = \theta^*$.

On aura alors $\mathbb{E}_{\theta^*}[Y_j] = M'(\theta^*) > 0$, de sorte que $P\{T_\ell < \infty\} = 1$ sous IS.

On sait que $M(\theta)$ est convexe, $M(0) = 1$ et $M'(0) = \mathbb{E}[Y_j] < 0$.

On suppose que $\theta^* \stackrel{\text{def}}{=} \sup\{\theta > 0 : M(\theta) \le 1\} < \infty$.



Pour IS, on va prendre $\theta = \theta^*$.

On aura alors $\mathbb{E}_{\theta^*}[Y_j] = M'(\theta^*) > 0$, de sorte que $P\{T_\ell < \infty\} = 1$ sous IS.
De plus,

$$L(\omega) \;=\; e^{-\theta^*\ell}e^{-\theta^*(D_{T_\ell}-\ell)} \;\le\; e^{-\theta^*\ell},$$

donc la variance est réduite au moins par le facteur $e^{-\theta^*\ell}$.

La probabilité à estimer s'écrit

$$\mu_\ell = e^{-\theta^* \ell}\, \mathbb{E}_{\theta^*}[\exp(-\theta^*(D_{T_\ell} - \ell))]$$

et l'estimateur IS estime cette dernière espérance.

La probabilité à estimer s'écrit

$$\mu_\ell = e^{-\theta^* \ell}\, \mathbb{E}_{\theta^*}[\exp(-\theta^*(D_{T_\ell} - \ell))]$$

et l'estimateur IS estime cette dernière espérance.

Cet exemple se généralise au cas où les lois des Y_j peuvent être différentes et pas toutes continues.

La probabilité à estimer s'écrit

$$\mu_\ell = e^{-\theta^*\ell}\, \mathbb{E}_{\theta^*}[\exp(-\theta^*(D_{T_\ell} - \ell))]$$

et l'estimateur IS estime cette dernière espérance.

Cet exemple se généralise au cas où les lois des Y_j peuvent être différentes et pas toutes continues.

En théorie du risque, $M(\theta) = 1$ s'appelle l'équation de Lundberg et θ^* est le paramètre de Lundberg.
Si ℓ est grand, on peut approximer μ_ℓ par $e^{-\theta^*\ell}$; c'est l'approximation de Lundberg. Ce que l'on vient de faire est exactement équivalent à utiliser cette approximation dans le schéma à variance zéro pour la chaîne de Markov.

Probabilité d'un très grand temps d'attente dans une file $GI/GI/1$.

A_j = temps entre les arrivées des clients j et $j+1$;
S_j = durée de service du client j;
W_j = durée d'attente du client j;

On suppose que A_j et S_j ont des densités h et g, $\mathbb{E}[S_j] < \mathbb{E}[A_j]$, et $W_1 = 0$.

On veut estimer

$$\mu_\ell = P[W > \ell].$$

Probabilité d'un très grand temps d'attente dans une file $GI/GI/1$.

A_j = temps entre les arrivées des clients j et $j+1$;
S_j = durée de service du client j;
W_j = durée d'attente du client j;

On suppose que A_j et S_j ont des densités h et g, $\mathbb{E}[S_j] < \mathbb{E}[A_j]$, et $W_1 = 0$.

On veut estimer

$$\mu_\ell = P[W > \ell].$$

Un théorème le la théorie des marches aléatoires nous dit que W (un temps d'attente à l'état stationnaire) suit la même loi que $\sup\{D_n,\, n \geq 0\}$ où $D_n = \sum_{j=1}^{n}(S_j - A_j)$. On a $\mu_\ell = \mathbb{P}[\sup\{D_n,\, n \geq 0\} > \ell]$.

Se ramène à l'exemple précédent en posant $Y_j = S_j - A_j$.

Probabilité d'un très grand temps d'attente dans une file $GI/GI/1$.

A_j = temps entre les arrivées des clients j et $j+1$;
S_j = durée de service du client j;
W_j = durée d'attente du client j;

On suppose que A_j et S_j ont des densités h et g, $\mathbb{E}[S_j] < \mathbb{E}[A_j]$, et $W_1 = 0$.

On veut estimer

$$\mu_\ell = P[W > \ell].$$

Un théorème le la théorie des marches aléatoires nous dit que W (un temps d'attente à l'état stationnaire) suit la même loi que $\sup\{D_n,\, n \geq 0\}$ où $D_n = \sum_{j=1}^{n}(S_j - A_j)$. On a $\mu_\ell = \mathbb{P}[\sup\{D_n,\, n \geq 0\} > \ell]$.

Se ramène à l'exemple précédent en posant $Y_j = S_j - A_j$.

La densité de Y_j est

$$\pi(y) = \int_0^\infty g(x)h(x-y)dx, \qquad y \in \mathbb{R},$$

et la densité

$$e^{\theta y}\pi(y) = \int_0^\infty e^{\theta x} g(x) e^{-\theta(x-y)} h(x-y)dx, \qquad y \in \mathbb{R}$$

s'obtient en appliquant la torsion exponentielle à h et g:

$$h_\theta(x) = e^{-\theta x} h(x)/M_h(-\theta)$$

et

$$g_\theta(x) = e^{\theta x} g(x)/M_g(\theta)$$

où

$$M_h(-\theta) = \int_0^\infty e^{-\theta x} h(x)dx = \mathbb{E}\left[e^{-\theta A_j}\right] < 1$$

et

$$M_g(\theta) = \int_0^\infty e^{\theta x} g(x)dx = \mathbb{E}\left[e^{\theta S_j}\right] > 1.$$

et la densité

$$e^{\theta y}\pi(y) = \int_0^\infty e^{\theta x} g(x) e^{-\theta(x-y)} h(x-y)dx, \qquad y \in \mathbb{R}$$

s'obtient en appliquant la torsion exponentielle à h et g:

$$h_\theta(x) = e^{-\theta x} h(x)/M_h(-\theta)$$

et

$$g_\theta(x) = e^{\theta x} g(x)/M_g(\theta)$$

où

$$M_h(-\theta) = \int_0^\infty e^{-\theta x} h(x)dx = \mathbb{E}\left[e^{-\theta A_j}\right] < 1$$

et

$$M_g(\theta) = \int_0^\infty e^{\theta x} g(x)dx = \mathbb{E}\left[e^{\theta S_j}\right] > 1.$$

Ici, M_h et M_g sont les fonction génératrices des moments de A_j et S_j, et $M(\theta) = M_g(\theta)M_h(-\theta)$ celle de Y_j.

On gonfle les durées de service et dégonfle les inter-arrivées.

On peut montrer que $\theta^* < \infty$ si $P[S_j > A_j] > 0$.

On gonfle les durées de service et dégonfle les inter-arrivées.

On peut montrer que $\theta^* < \infty$ si $P[S_j > A_j] > 0$.

En prenant $\theta = \theta^*$, on obtient

$$\mathbb{E}_{\theta^*}[S_j - A_j] = M'(\theta^*) > 0,$$

$P[T_\ell < \infty] = 1$ et

$$L(\omega) \ = \ e^{-\theta^*\ell} e^{-\theta^*(D_{T_\ell} - \ell)} \ \leq \ e^{-\theta^*\ell}$$

lorsque $T_\ell < \infty$.

Temps jusqu'au premier débordement dans une file $GI/GI/1/K$.

T_K = premier instant où il y a K clients dans le système. $\mu = \mathbb{E}[T_k]$.

Temps jusqu'au premier débordement dans une file $GI/GI/1/K$.

$T_K =$ premier instant où il y a K clients dans le système. $\mu = \mathbb{E}[T_k]$.

E.g.: débordement d'un tampon dans un commutateur.

Temps jusqu'au premier débordement dans une file $GI/GI/1/K$.

T_K = premier instant où il y a K clients dans le système. $\mu = \mathbb{E}[T_k]$.

E.g.: débordement d'un tampon dans un commutateur.

Ce système possède un point de regénération au temps T_0 où un premier client arrive après que le système se soit vidé.

Temps jusqu'au premier débordement dans une file $GI/GI/1/K$.

$T_K =$ premier instant où il y a K clients dans le système. $\mu = \mathbb{E}[T_k]$.

E.g.: débordement d'un tampon dans un commutateur.

Ce système possède un point de regénération au temps T_0 où un premier client arrive après que le système se soit vidé.

Soit $T = \min(T_0, T_K)$.

Temps jusqu'au premier débordement dans une file $GI/GI/1/K$.

T_K = premier instant où il y a K clients dans le système. $\mu = \mathbb{E}[T_k]$.

E.g.: débordement d'un tampon dans un commutateur.

Ce système possède un point de regénération au temps T_0 où un premier client arrive après que le système se soit vidé.

Soit $T = \min(T_0, T_K)$.

On a $T_K = T + (T_K - T)\mathbb{I}[T = T_0]$ et

$$\begin{aligned} \mathbb{E}[T_K] &= \mathbb{E}[T] + \mathbb{E}[T_K - T \mid T = T_0] \cdot \mathbb{P}[T = T_0] \\ &= \mathbb{E}[T] + \mathbb{E}[T_K] P[T = T_0], \end{aligned}$$

Temps jusqu'au premier débordement dans une file $GI/GI/1/K$.

T_K = premier instant où il y a K clients dans le système. $\mu = \mathbb{E}[T_k]$.

E.g.: débordement d'un tampon dans un commutateur.

Ce système possède un point de regénération au temps T_0 où un premier client arrive après que le système se soit vidé.

Soit $T = \min(T_0, T_K)$.

On a $T_K = T + (T_K - T)\mathbb{I}[T = T_0]$ et

$$\begin{aligned}
\mathbb{E}[T_K] &= \mathbb{E}[T] + \mathbb{E}[T_K - T \mid T = T_0] \cdot \mathbb{P}[T = T_0] \\
&= \mathbb{E}[T] + \mathbb{E}[T_K]P[T = T_0],
\end{aligned}$$

qui se réécrit

$$\mathbb{E}[T_K] = \frac{\mathbb{E}[T]}{(1 - \mathbb{P}[T = T_0])} = \frac{\mathbb{E}[T]}{\mathbb{P}[T = T_K]}.$$

Temps jusqu'au premier débordement dans une file $GI/GI/1/K$.

T_K = premier instant où il y a K clients dans le système. $\mu = \mathbb{E}[T_k]$.

E.g.: débordement d'un tampon dans un commutateur.

Ce système possède un point de regénération au temps T_0 où un premier client arrive après que le système se soit vidé.

Soit $T = \min(T_0, T_K)$.

On a $T_K = T + (T_K - T)\mathbb{I}[T = T_0]$ et

$$\begin{aligned}\mathbb{E}[T_K] &= \mathbb{E}[T] + \mathbb{E}[T_K - T \mid T = T_0] \cdot \mathbb{P}[T = T_0] \\ &= \mathbb{E}[T] + \mathbb{E}[T_K] P[T = T_0],\end{aligned}$$

qui se réécrit

$$\mathbb{E}[T_K] = \frac{\mathbb{E}[T]}{(1 - \mathbb{P}[T = T_0])} = \frac{\mathbb{E}[T]}{\mathbb{P}[T = T_K]}.$$

Si le débordement est rare, $p_K = \mathbb{P}[T = T_K]$ est difficile à estimer, mais $\mathbb{E}[T]$ est habituellement facile à estimer.

On estimera alors $\mathbb{E}[T]$ par MC ordinaire et p_K par IS, avec la même torsion exponentielle que dans l'exemple précédent.

L'estimateur IS de p_K est $X_{\rm is} = \mathbb{I}[T = T_K] L(\omega)$.

On estimera alors $\mathbb{E}[T]$ par MC ordinaire et p_K par IS, avec la même torsion exponentielle que dans l'exemple précédent.

L'estimateur IS de p_K est $X_{\rm is} = \mathbb{I}[T = T_K] L(\omega)$.

Si $T = T_K$, soient N_A et N_S les nombres de A_j et S_j générés durant $[0, T)$.

Au temps T_K, il y a K clients dans le système, le client $(N_A + 1)$ arrive, et $N_S - 1$ clients ont quitté.

On estimera alors $\mathbb{E}[T]$ par MC ordinaire et p_K par IS, avec la même torsion exponentielle que dans l'exemple précédent.

L'estimateur IS de p_K est $X_{\text{is}} = \mathbb{I}[T = T_K] L(\omega)$.

Si $T = T_K$, soient N_A et N_S les nombres de A_j et S_j générés durant $[0, T)$.

Au temps T_K, il y a K clients dans le système, le client $(N_A + 1)$ arrive, et $N_S - 1$ clients ont quitté. On a donc

$$N_A = N_S + K - 2,$$

On estimera alors $\mathbb{E}[T]$ par MC ordinaire et p_K par IS, avec la même torsion exponentielle que dans l'exemple précédent.

L'estimateur IS de p_K est $X_{\text{is}} = \mathbb{I}[T = T_K]L(\omega)$.

Si $T = T_K$, soient N_A et N_S les nombres de A_j et S_j générés durant $[0, T)$.

Au temps T_K, il y a K clients dans le système, le client $(N_A + 1)$ arrive, et $N_S - 1$ clients ont quitté. On a donc

$$
\begin{aligned}
N_A &= N_S + K - 2, \\
T_K &= \sum_{j=1}^{N_A} A_j \leq \sum_{j=1}^{N_S} S_j,
\end{aligned}
$$

On estimera alors $\mathbb{E}[T]$ par MC ordinaire et p_K par IS, avec la même torsion exponentielle que dans l'exemple précédent.

L'estimateur IS de p_K est $X_{\text{is}} = \mathbb{I}[T = T_K]L(\omega)$.

Si $T = T_K$, soient N_A et N_S les nombres de A_j et S_j générés durant $[0, T)$.

Au temps T_K, il y a K clients dans le système, le client $(N_A + 1)$ arrive, et $N_S - 1$ clients ont quitté. On a donc

$$
\begin{aligned}
N_A &= N_S + K - 2, \\
T_K &= \sum_{j=1}^{N_A} A_j \leq \sum_{j=1}^{N_S} S_j, \\
L(\omega) &= \prod_{j=1}^{N_A} \frac{h(A_j)}{h_\theta(A_j)} \prod_{j=1}^{N_S} \frac{g(S_j)}{g_\theta(S_j)} \\
&= [M_h(-\theta)]^{K-2}[M_h(-\theta)M_g(\theta)]^{N_S} \exp\left(\theta \sum_{j=1}^{N_A} A_j - \theta \sum_{j=1}^{N_S} S_j\right)
\end{aligned}
$$

$$\leq \quad [M_h(-\theta)]^{K-2}[M(\theta)]^{N_S}.$$

Avec $\theta = \theta^*$, on obtient

$$L(\omega) \quad = \quad M_h(-\theta^*)^{K-2} \exp\left(\theta \sum_{j=1}^{N_A} A_j - \theta \sum_{j=1}^{N_S} S_j\right) \;\leq\; M_h(-\theta^*)^{K-2} \;<\; 1.$$

Cas particulier: file $M/M/1$.

Cas particulier: file $M/M/1$.

Taux d'arrivée λ, taux de service μ, $\lambda < \mu$.

Cas particulier: file $M/M/1$.

Taux d'arrivée λ, taux de service μ, $\lambda < \mu$.

La densité de A_j est $h(y) = \lambda \exp(-\lambda y)$ pour $y > 0$ et on a

$$M_h(-\theta) = \int_0^\infty e^{-\theta y} \lambda e^{-\lambda y} dy = \frac{\lambda}{\lambda + \theta}$$

pour $\theta > -\lambda$.

Cas particulier: file $M/M/1$.

Taux d'arrivée λ, taux de service μ, $\lambda < \mu$.

La densité de A_j est $h(y) = \lambda \exp(-\lambda y)$ pour $y > 0$ et on a

$$M_h(-\theta) = \int_0^\infty e^{-\theta y} \lambda e^{-\lambda y} dy = \frac{\lambda}{\lambda + \theta}$$

pour $\theta > -\lambda$. De même, pour les durées de service,

$$M_g(\theta) = \frac{\mu}{(\mu - \theta)} \quad \text{pour } \theta < \mu.$$

Cas particulier: file $M/M/1$.

Taux d'arrivée λ, taux de service μ, $\lambda < \mu$.

La densité de A_j est $h(y) = \lambda \exp(-\lambda y)$ pour $y > 0$ et on a

$$M_h(-\theta) = \int_0^\infty e^{-\theta y} \lambda e^{-\lambda y} dy = \frac{\lambda}{\lambda + \theta}$$

pour $\theta > -\lambda$. De même, pour les durées de service,

$$M_g(\theta) = \frac{\mu}{(\mu - \theta)} \quad \text{pour } \theta < \mu.$$

L'équation $M(\theta) = M_h(-\theta)M_g(\theta) = 1$ devient $\lambda\mu = (\lambda + \theta)(\mu - \theta)$ et la solution est $\theta^* = \mu - \lambda$.

Cas particulier: file $M/M/1$.

Taux d'arrivée λ, taux de service μ, $\lambda < \mu$.

La densité de A_j est $h(y) = \lambda \exp(-\lambda y)$ pour $y > 0$ et on a

$$M_h(-\theta) = \int_0^\infty e^{-\theta y} \lambda e^{-\lambda y} dy = \frac{\lambda}{\lambda + \theta}$$

pour $\theta > -\lambda$. De même, pour les durées de service,

$$M_g(\theta) = \frac{\mu}{(\mu - \theta)} \quad \text{pour } \theta < \mu.$$

L'équation $M(\theta) = M_h(-\theta) M_g(\theta) = 1$ devient $\lambda\mu = (\lambda + \theta)(\mu - \theta)$ et la solution est $\theta^* = \mu - \lambda$.

Les nouvelles densités avec $\theta = \theta^*$ sont

$$h_\theta(x) = \frac{e^{-\theta^* x} h(x)}{M_h(-\theta^*)}$$

Cas particulier: file $M/M/1$.

Taux d'arrivée λ, taux de service μ, $\lambda < \mu$.

La densité de A_j est $h(y) = \lambda \exp(-\lambda y)$ pour $y > 0$ et on a

$$M_h(-\theta) = \int_0^\infty e^{-\theta y} \lambda e^{-\lambda y} dy = \frac{\lambda}{\lambda + \theta}$$

pour $\theta > -\lambda$. De même, pour les durées de service,

$$M_g(\theta) = \frac{\mu}{(\mu - \theta)} \quad \text{pour } \theta < \mu.$$

L'équation $M(\theta) = M_h(-\theta) M_g(\theta) = 1$ devient $\lambda\mu = (\lambda + \theta)(\mu - \theta)$ et la solution est $\theta^* = \mu - \lambda$.

Les nouvelles densités avec $\theta = \theta^*$ sont

$$h_\theta(x) = \frac{e^{-\theta^* x} h(x)}{M_h(-\theta^*)} = \frac{e^{-(\mu-\lambda)x} \lambda e^{-\lambda x}}{M_h(-\theta^*)}$$

Cas particulier: file $M/M/1$.

Taux d'arrivée λ, taux de service μ, $\lambda < \mu$.

La densité de A_j est $h(y) = \lambda \exp(-\lambda y)$ pour $y > 0$ et on a

$$M_h(-\theta) = \int_0^\infty e^{-\theta y} \lambda e^{-\lambda y} dy = \frac{\lambda}{\lambda + \theta}$$

pour $\theta > -\lambda$. De même, pour les durées de service,

$$M_g(\theta) = \frac{\mu}{(\mu - \theta)} \quad \text{pour } \theta < \mu.$$

L'équation $M(\theta) = M_h(-\theta) M_g(\theta) = 1$ devient $\lambda\mu = (\lambda + \theta)(\mu - \theta)$ et la solution est $\theta^* = \mu - \lambda$.

Les nouvelles densités avec $\theta = \theta^*$ sont

$$h_\theta(x) = \frac{e^{-\theta^* x} h(x)}{M_h(-\theta^*)} = \frac{e^{-(\mu-\lambda)x} \lambda e^{-\lambda x}}{M_h(-\theta^*)} = \mu e^{-\mu x}$$

et

$$g_\theta(x) = \frac{e^{\theta x} g(x)}{M_g(\theta^*)}$$

et

$$g_\theta(x) = \frac{e^{\theta x} g(x)}{M_g(\theta^*)} = \frac{e^{(\mu-\lambda)x} \lambda e^{-\mu x}}{M_g(\theta^*)} = \lambda e^{-\lambda x}.$$

et

$$g_\theta(x) = \frac{e^{\theta x} g(x)}{M_g(\theta^*)} = \frac{e^{(\mu-\lambda)x} \lambda e^{-\mu x}}{M_g(\theta^*)} = \lambda e^{-\lambda x}.$$

IS permute simplement μ et λ.

et

$$g_\theta(x) = \frac{e^{\theta x} g(x)}{M_g(\theta^*)} = \frac{e^{(\mu-\lambda)x} \lambda e^{-\mu x}}{M_g(\theta^*)} = \lambda e^{-\lambda x}.$$

IS permute simplement μ et λ.

Puisque $M_h(-\theta^*) = \lambda/\mu$,
la variance est divisée par au moins $e^{(\mu-\lambda)\ell}$ lorsqu'on estime $\mu_\ell = P[W > \ell]$

et

$$g_\theta(x) = \frac{e^{\theta x} g(x)}{M_g(\theta^*)} = \frac{e^{(\mu-\lambda)x}\lambda e^{-\mu x}}{M_g(\theta^*)} = \lambda e^{-\lambda x}.$$

IS permute simplement μ et λ.

Puisque $M_h(-\theta^*) = \lambda/\mu$,
la variance est divisée par au moins $e^{(\mu-\lambda)\ell}$ lorsqu'on estime $\mu_\ell = P[W > \ell]$
et par $(\mu/\lambda)^{K-2}$ lorsqu'on estime $p_K = P[T = T_K]$.

Probabilité de ruine d'une compagnie d'assurance.

La compagnie encaisse des primes au taux $c > 0$.
Les réclamations arrivent selon un processus de Poisson $\{N(t),\, t \geq 0\}$ de taux λ et leurs tailles sont des v.a. i.i.d. $C_1, C_2, \ldots$ de densité h telle que $M_h(\theta) < \infty$ autour de 0.

Probabilité de ruine d'une compagnie d'assurance.

La compagnie encaisse des primes au taux $c > 0$.
Les réclamations arrivent selon un processus de Poisson $\{N(t),\, t \geq 0\}$ de taux λ et leurs tailles sont des v.a. i.i.d. $C_1, C_2, \ldots$ de densité h telle que $M_h(\theta) < \infty$ autour de 0. La réserve au temps t est

$$R(t) = R(0) + ct - \sum_{j=1}^{N(t)} C_j.$$

On veut estimer $\mu = \mathbb{P}[\inf_{t>0} R(t) < 0]$, la probabilité de ruine.

Probabilité de ruine d'une compagnie d'assurance.

La compagnie encaisse des primes au taux $c > 0$.
Les réclamations arrivent selon un processus de Poisson $\{N(t),\, t \geq 0\}$ de taux λ et leurs tailles sont des v.a. i.i.d. $C_1, C_2, \ldots$ de densité h telle que $M_h(\theta) < \infty$ autour de 0. La réserve au temps t est

$$R(t) = R(0) + ct - \sum_{j=1}^{N(t)} C_j.$$

On veut estimer $\mu = \mathbb{P}[\inf_{t>0} R(t) < 0]$, la probabilité de ruine.

En écrivant

$$R(0) - R(t) = \sum_{j=1}^{N(t)} (C_j - A_j c) = \sum_{j=1}^{N(t)} Y_j = D_{N(t)}$$

où $Y_j = C_j - A_j c$ et A_j la durée entre les réclamations $j-1$ et j, on se ramène à

$$\mu = \mu_\ell = \mathbb{P}\left[\sup_{t\geq 0} D_{N(t)} > \ell \stackrel{\text{def}}{=} R(0)\right] = \mathbb{P}[T_\ell < \infty].$$

La fonction génératrice des A_j est

$$M_a(\theta) = \int_0^\infty e^{\theta x} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - \theta},$$

et celle des Y_j est donc

$$M(\theta) = \mathbb{E}\left[e^{\theta(C_j - cA_j)}\right] = M_h(\theta) M_a(-\theta c) = M_h(\theta) \frac{\lambda}{\lambda + \theta c}.$$

La fonction génératrice des A_j est

$$M_a(\theta) = \int_0^\infty e^{\theta x} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - \theta},$$

et celle des Y_j est donc

$$M(\theta) = \mathbb{E}\left[e^{\theta(C_j - cA_j)}\right] = M_h(\theta) M_a(-\theta c) = M_h(\theta) \frac{\lambda}{\lambda + \theta c}.$$

IS remplace $h(x)$ par $h_\theta(x) = h(x) e^{\theta x}/M_h(\theta)$ et la densité exponentielle des A_j par $(\lambda + \theta c) e^{-(\lambda + \theta c)x}$ (i.e., augmente λ à $\lambda_\theta = \lambda + \theta c$).

L'équation de Lundberg, $M(\theta) = 1$, s'écrit $M_h(\theta) = (\lambda + \theta c)/\lambda$, et $\theta^* > 0$ est la plus grande solution de cette équation.

Avec $\theta = \theta^*$, on obtient

$$\lambda_{\theta^*} = \lambda M_h(\theta^*),$$

$$\mathbb{E}_{\theta^*}[C_j] = M_h'(\theta^*) = c/\lambda,$$

de sorte que

$$\mathbb{E}_{\theta^*}[R(0) - R(t)] = \lambda t \, \mathbb{E}_{\theta^*}[C_j] - ct = 0$$

et la ruine survient avec probabilité 1.

Avec $\theta = \theta^*$, on obtient

$$\lambda_{\theta^*} = \lambda M_h(\theta^*),$$
$$\mathbb{E}_{\theta^*}[C_j] = M_h'(\theta^*) = c/\lambda,$$

de sorte que

$$\mathbb{E}_{\theta^*}[R(0) - R(t)] = \lambda t \, \mathbb{E}_{\theta^*}[C_j] - ct = 0$$

et la ruine survient avec probabilité 1.

Le rapport de vraisemblance à l'instant de ruine T_ℓ est

$$L(\omega) = e^{\theta^*(R(T_\ell) - R(0))} \leq e^{-\theta^* R(0)}.$$

Voir chap. 1 pour un exemple numérique.