Apprentissage automatique

Méthodes d'échantillonnage - motivation

ÉCHANTILLONNAGE

Sujets: échantillonnage

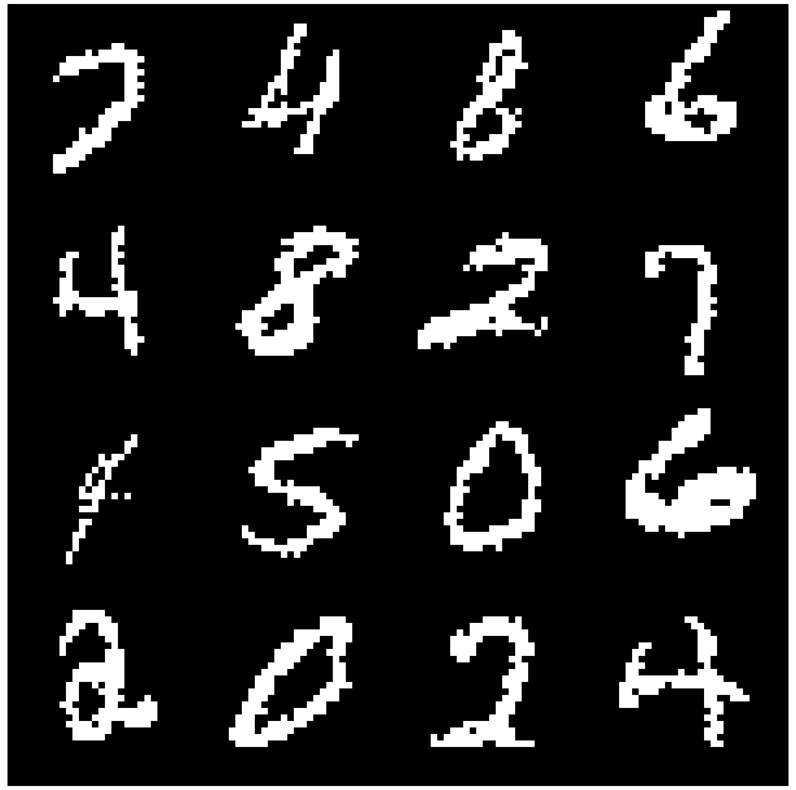
- La plupart des algorithmes d'apprentissage vus sont basés sur un modèle probabiliste
  - un modèle probabiliste exprime comment on suppose que nos données ont été générées
- Comment simule-t-on ce processus de génération ?
  - on utilise un algorithme d'**échantillonnage** tel que :

  probabilité de générer une donnée = probabilité assignée par le modèle

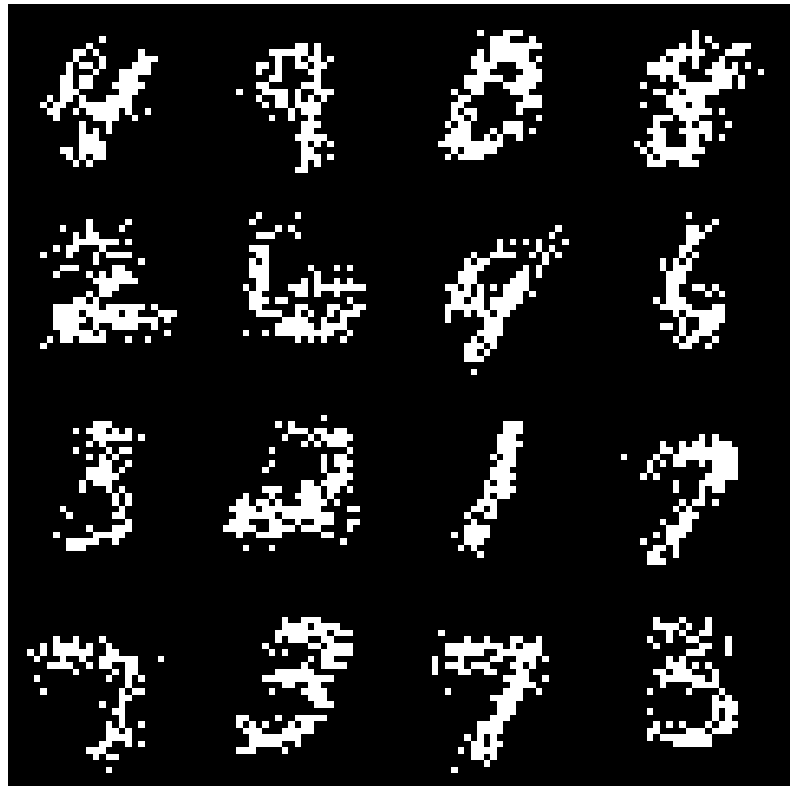
ÉCHANTILLONNAGE

Sujets: visualisation

- Pourquoi échantillonner ?
  - pour visualiser ce qui a été appris



Ensemble d'entraînement



Mélange de produits de Bernoulli

Salakhutdinov et Murray, 2008

ÉCHANTILLONNAGE

Sujets: estimation d'espérance, estimation Monte Carlo

- Pourquoi échantillonner ?
  - pour estimer une espérance

$$\mathbb{E}[f] = \int f(\mathbf{z})p(\mathbf{z})\,\mathrm{d}\mathbf{z}$$

ÉCHANTILLONNAGE

Sujets: estimation d'espérance, estimation Monte Carlo

- Pourquoi échantillonner ?
  - pour estimer une espérance

$$\mathbb{E}[f] = \int f(\mathbf{z})p(\mathbf{z})\,\mathrm{d}\mathbf{z} \;\approx\; \frac{1}{L}\sum_{l=1}^{L} f(\mathbf{z}^{(l)})$$

$\mathbf{z}^{(l)}$ échantillonnés de $p(\mathbf{z})$

- Un tel calcul d'une espérance ainsi est appelé une **estimation Monte Carlo**

ÉCHANTILLONNAGE

Sujets: estimation de la loi prédictive a posteriori

- Pourquoi échantillonner ?
  - pour estimer la loi prédictive a posteriori

$$p(t|\mathbf{t}, \alpha, \beta) = \int p(t|\mathbf{w}, \beta) p(\mathbf{w}|\mathbf{t}, \alpha, \beta)\, \mathrm{d}\mathbf{w}$$

ÉCHANTILLONNAGE

Sujets: estimation de la loi prédictive a posteriori

- Pourquoi échantillonner ?
  - pour estimer la loi prédictive a posteriori

$$p(t|\mathbf{t}, \alpha, \beta) = \int p(t|\mathbf{w}, \beta) p(\mathbf{w}|\mathbf{t}, \alpha, \beta)\, \mathrm{d}\mathbf{w}$$

$$\approx \frac{1}{L} \sum_{l=1}^{L} p(t|\mathbf{w}^{(l)}, \beta)$$

$\mathbf{w}^{(l)}$ échantillonnés de $p(\mathbf{w}|\mathbf{t}, \alpha, \beta)$

ÉCHANTILLONNAGE

Sujets: estimation de la loi prédictive a posteriori

- Pourquoi échantillonner ?
    - pour approximer l'étape M de l'algorithme EM

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{old}}) = \int p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\mathrm{old}}) \ln p(\mathbf{Z}, \mathbf{X}|\boldsymbol{\theta}) \, \mathrm{d}\mathbf{Z}$$

ÉCHANTILLONNAGE

Sujets: estimation de la loi prédictive a posteriori

- Pourquoi échantillonner ?
  - pour approximer l'étape M de l'algorithme EM

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \int p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{Z}, \mathbf{X}|\boldsymbol{\theta}) \, \mathrm{d}\mathbf{Z}$$

$$\simeq \frac{1}{L} \sum_{l=1}^{L} \ln p(\mathbf{Z}^{(l)}, \mathbf{X}|\boldsymbol{\theta})$$

$\mathbf{Z}^{(l)}$ échantillonnés de $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$

Apprentissage automatique

Méthodes d’échantillonnage - variable aléatoire discrète scalaire

ÉCHANTILLONNAGE DE BASE

Sujets: méthodes de base

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire
  - d'une variable aléatoire **continue** scalaire
  - d'une variable aléatoire **gaussienne vectorielle**

- On va supposer qu'on a accès à un générateur pour une variable aléatoire uniformément distribuée entre 0 et 1
  - les méthodes de base vont utiliser les nombres aléatoires qu'il génère

ÉCHANTILLONNAGE DE BASE

Sujets: méthodes de base

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire
  - d'une variable aléatoire **continue** scalaire
  - d'une variable aléatoire **gaussienne vectorielle**

- On va supposer qu'on a accès à un générateur pour une variable aléatoire uniformément distribuée entre 0 et 1
  - les méthodes de base vont utiliser les nombres aléatoires qu'il génère

ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire discrète scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire

- Variable : $Z \in \{1,2,\ldots,M\}$

- Algorithme

1. séparer l'intervalle $(0,1)$ en M intervalles, où le i^{e} intervalle est de longueur $p(Z=i)$ et est associé à la valeur i

2. échantillonner un nombre x uniformément dans $(0,1)$

3. retourner l'étiquette de l'intervalle dans lequel x se trouve

ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire discrète scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire

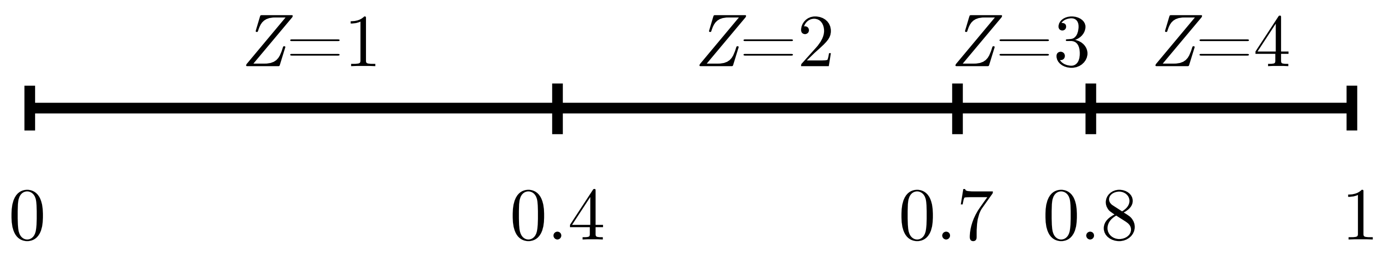
$$
\begin{aligned}
p(Z{=}1) &= 0.4 \\
p(Z{=}2) &= 0.3 \\
p(Z{=}3) &= 0.1 \\
p(Z{=}4) &= 0.2
\end{aligned}
$$

ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire discrète scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire

$$
\begin{aligned}
p(Z=1) &= 0.4 \\
p(Z=2) &= 0.3 \\
p(Z=3) &= 0.1 \\
p(Z=4) &= 0.2
\end{aligned}
$$



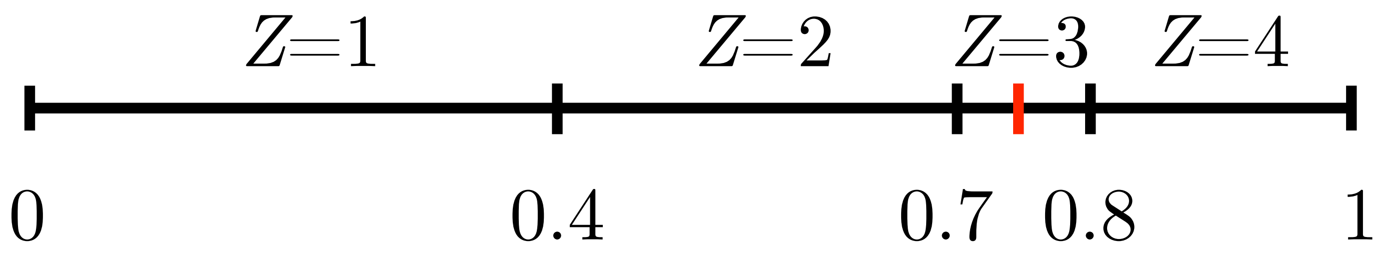
ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire discrète scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire

$$
\begin{aligned}
p(Z{=}1) &= 0.4\\
p(Z{=}2) &= 0.3\\
p(Z{=}3) &= 0.1\\
p(Z{=}4) &= 0.2
\end{aligned}
$$

$$x = 0.7456$$



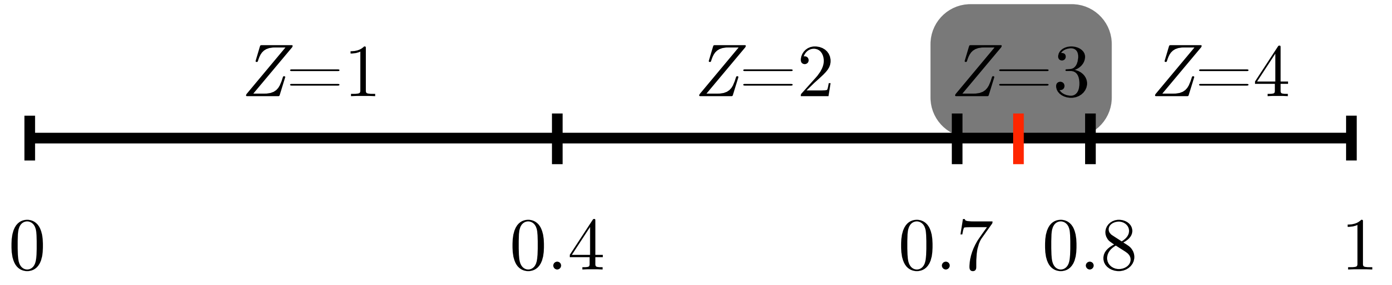
Échantillonnage de base

Sujets: variable aléatoire discrète scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire

$$
\begin{aligned}
p(Z=1) &= 0.4 \\
p(Z=2) &= 0.3 \\
p(Z=3) &= 0.1 \\
p(Z=4) &= 0.2
\end{aligned}
$$

$x = 0.7456$



ÉCHANTILLONNAGE DE BASE

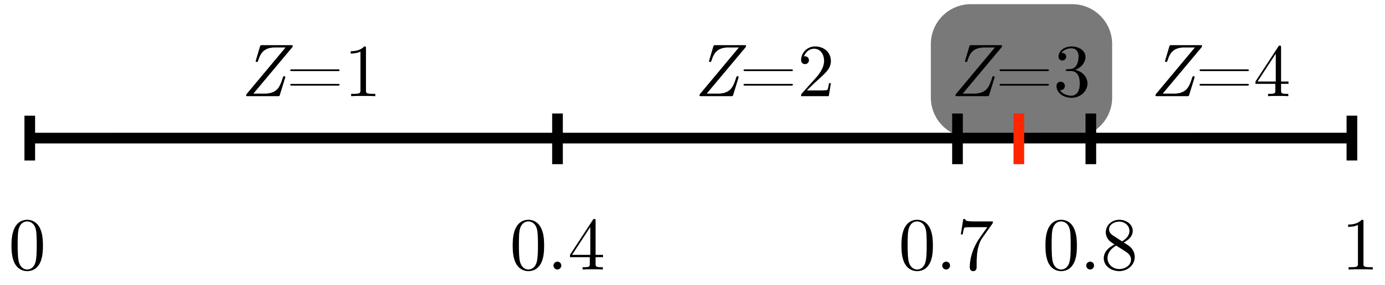
Sujets: variable aléatoire discrète scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire

$$\begin{aligned} p(Z{=}1) &= 0.4 \\ p(Z{=}2) &= 0.3 \\ p(Z{=}3) &= 0.1 \\ p(Z{=}4) &= 0.2 \end{aligned}$$

$$p(0.7 < x < 0.8)$$

$$x = 0.7456$$



ÉCHANTILLONNAGE DE BASE

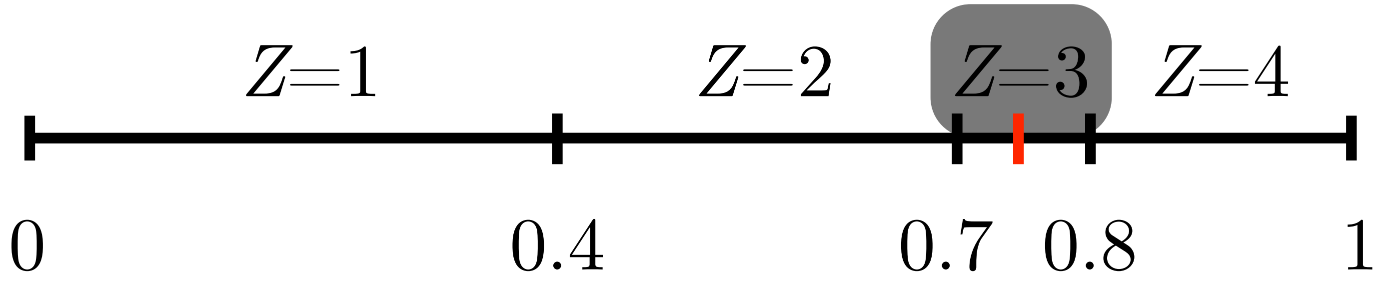
Sujets: variable aléatoire discrète scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire

$$
\begin{aligned}
p(Z{=}1) &= 0.4 \\
p(Z{=}2) &= 0.3 \\
p(Z{=}3) &= 0.1 \\
p(Z{=}4) &= 0.2
\end{aligned}
$$

$$\overbrace{p(0.7 < x < 0.8)}^{\text{retourne } 3}$$

$$x = 0.7456$$



ÉCHANTILLONNAGE DE BASE

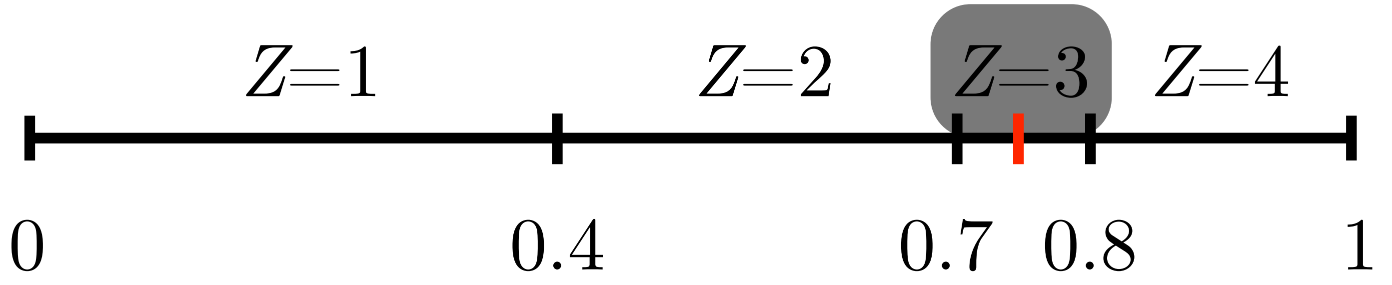
Sujets: variable aléatoire discrète scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire

$$
\begin{aligned}
p(Z{=}1) &= 0.4\\
p(Z{=}2) &= 0.3\\
p(Z{=}3) &= 0.1\\
p(Z{=}4) &= 0.2
\end{aligned}
$$

$$\overbrace{p(0.7 < x < 0.8)}^{\text{retourne } 3} = \int_{0.7}^{0.8} 1 \mathrm{d}x = 0.1$$

$$x = 0.7456$$



Apprentissage automatique

Méthodes d’échantillonnage - variable aléatoire continue scalaire

ÉCHANTILLONNAGE DE BASE

Sujets: méthodes de base

RAPPEL

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire
  - d'une variable aléatoire **continue** scalaire
  - d'une variable aléatoire **gaussienne vectorielle**

- On va supposer qu'on a accès à un générateur pour une variable aléatoire uniformément distribuée entre 0 et 1
  - les méthodes de base vont utiliser les nombres aléatoires qu'il génère

ÉCHANTILLONNAGE DE BASE

Sujets: méthodes de base

RAPPEL

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire
  - d'une variable aléatoire **continue** scalaire
  - d'une variable aléatoire **gaussienne vectorielle**

- On va supposer qu'on a accès à un générateur pour une variable aléatoire uniformément distribuée entre 0 et 1
  - les méthodes de base vont utiliser les nombres aléatoires qu'il génère

ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire continue scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **continue** scalaire
- Variable : $Z \in \mathrm{I}$, fonction de densité $p(z)$
  et fonction de répartition $P(z) = \int_{\min(I)}^{z} p(x) \mathrm{d}x$
- Algorithme

1. échantillonner un nombre x uniformément dans $(0,1)$
2. retourner $P^{-1}(x)$

ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire continue scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **continue** scalaire
- Variable : $Z \in \mathrm{I}$, fonction de densité $p(z)$

  et fonction de répartition $P(z) = \int_{\min(I)}^{z} p(x)\mathrm{d}x$
- Algorithme

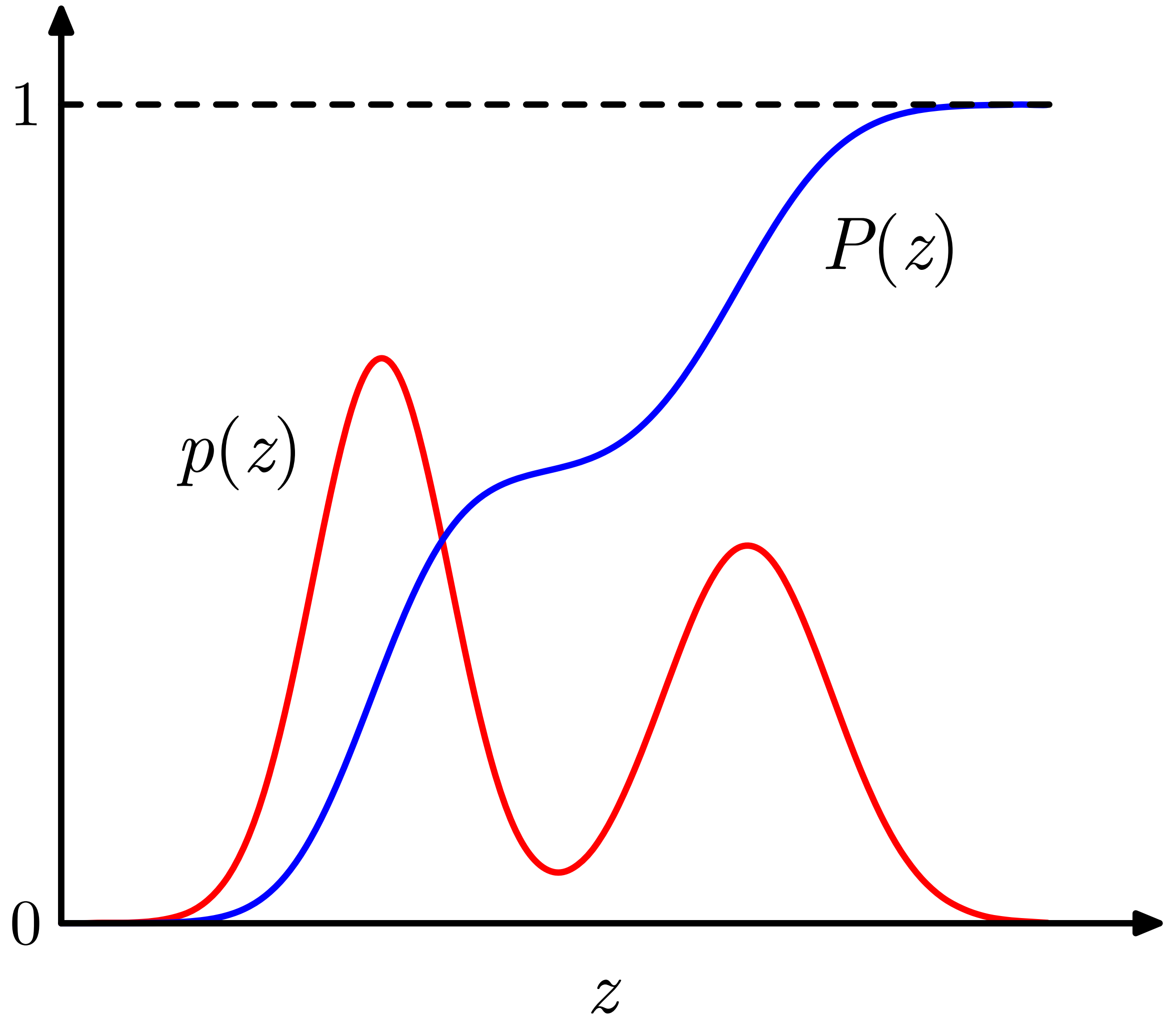
  1. échantillonner un nombre x uniformément dans $(0{,}1)$
  2. retourner $P^{-1}(x)$

  fonction réciproque de $P(z)$ (*inverse function*)

ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire continue scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **continue** scalaire

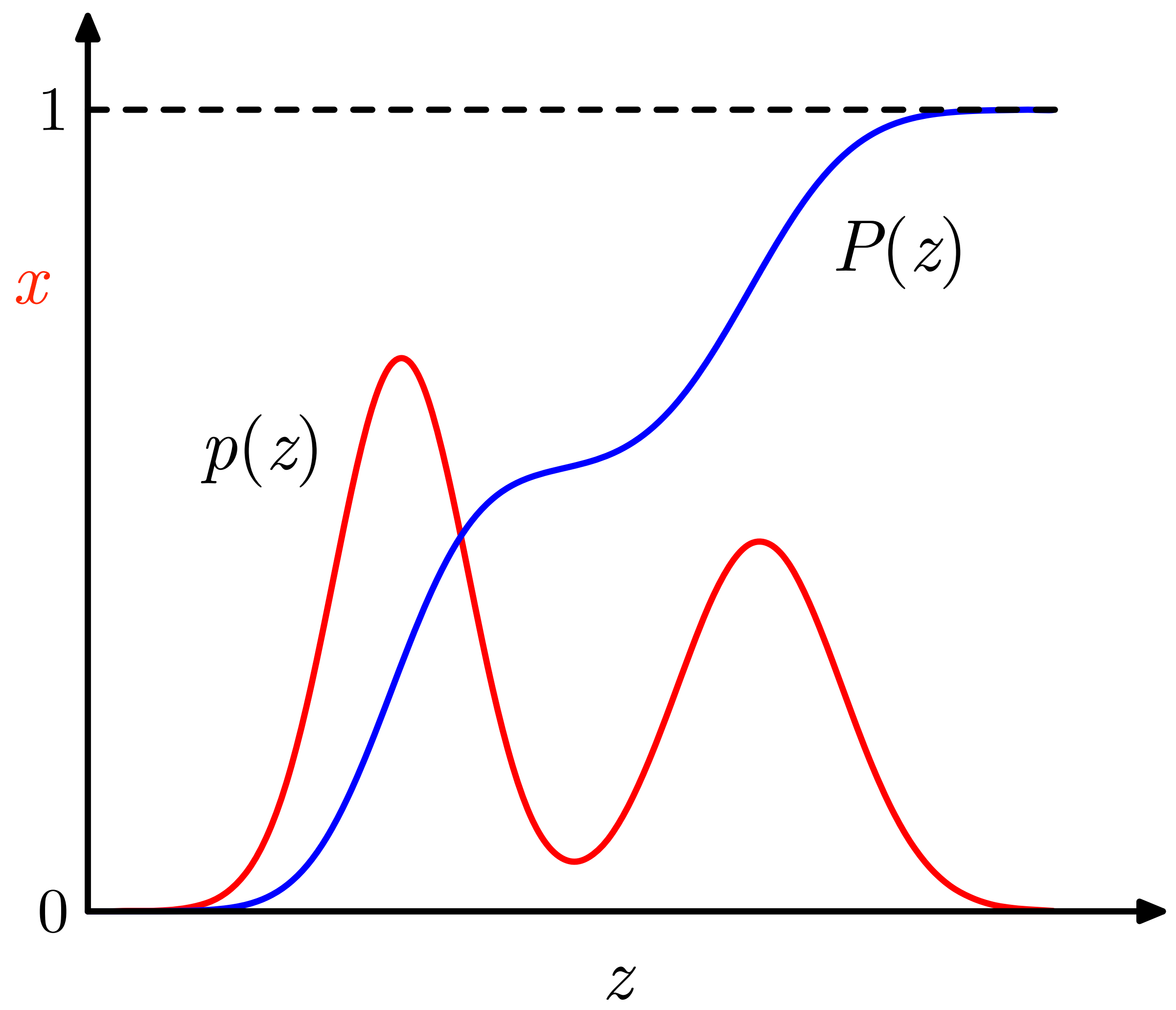


HUGO LAROCHELLE

ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire continue scalaire

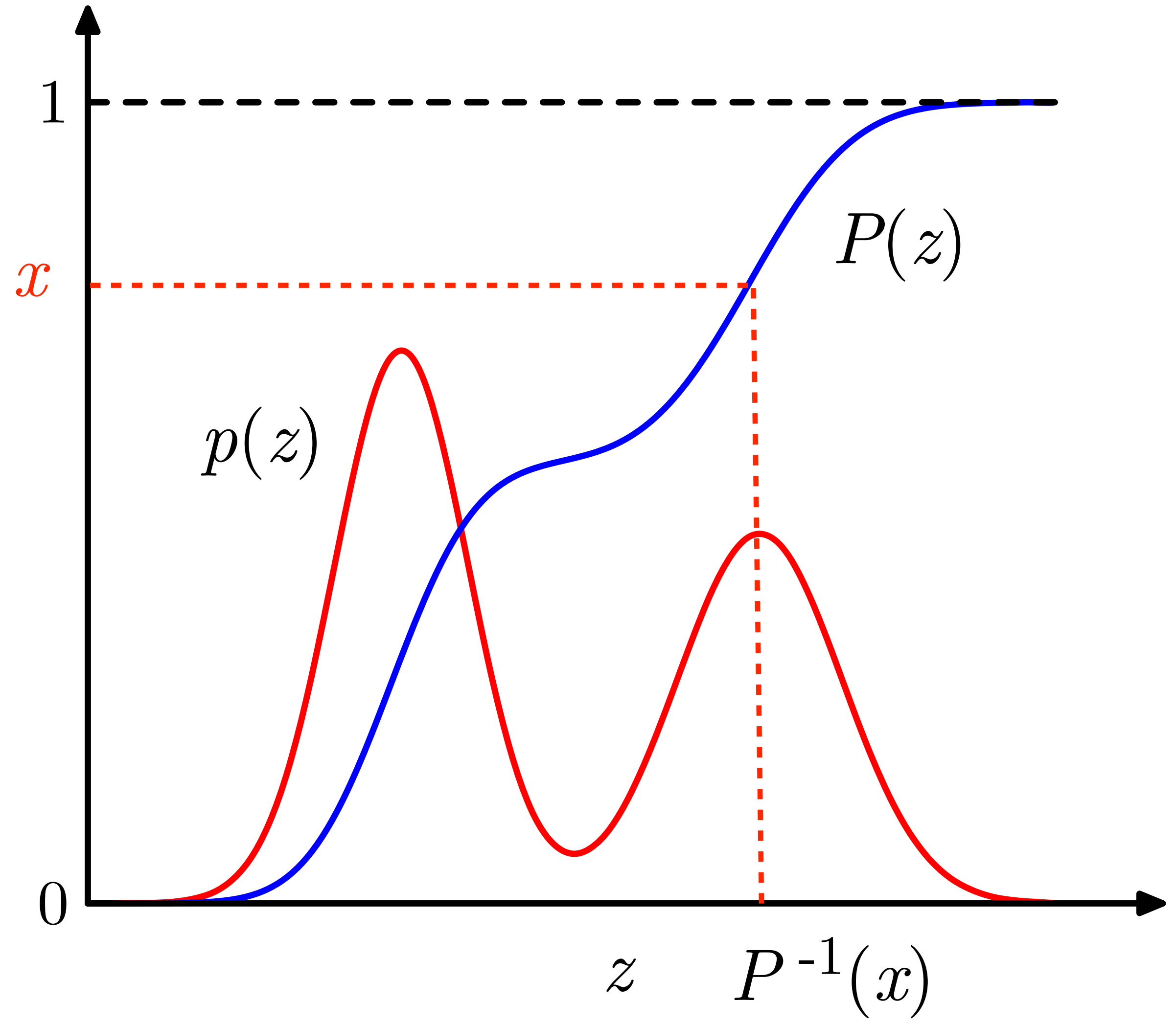
- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **continue** scalaire



ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire continue scalaire

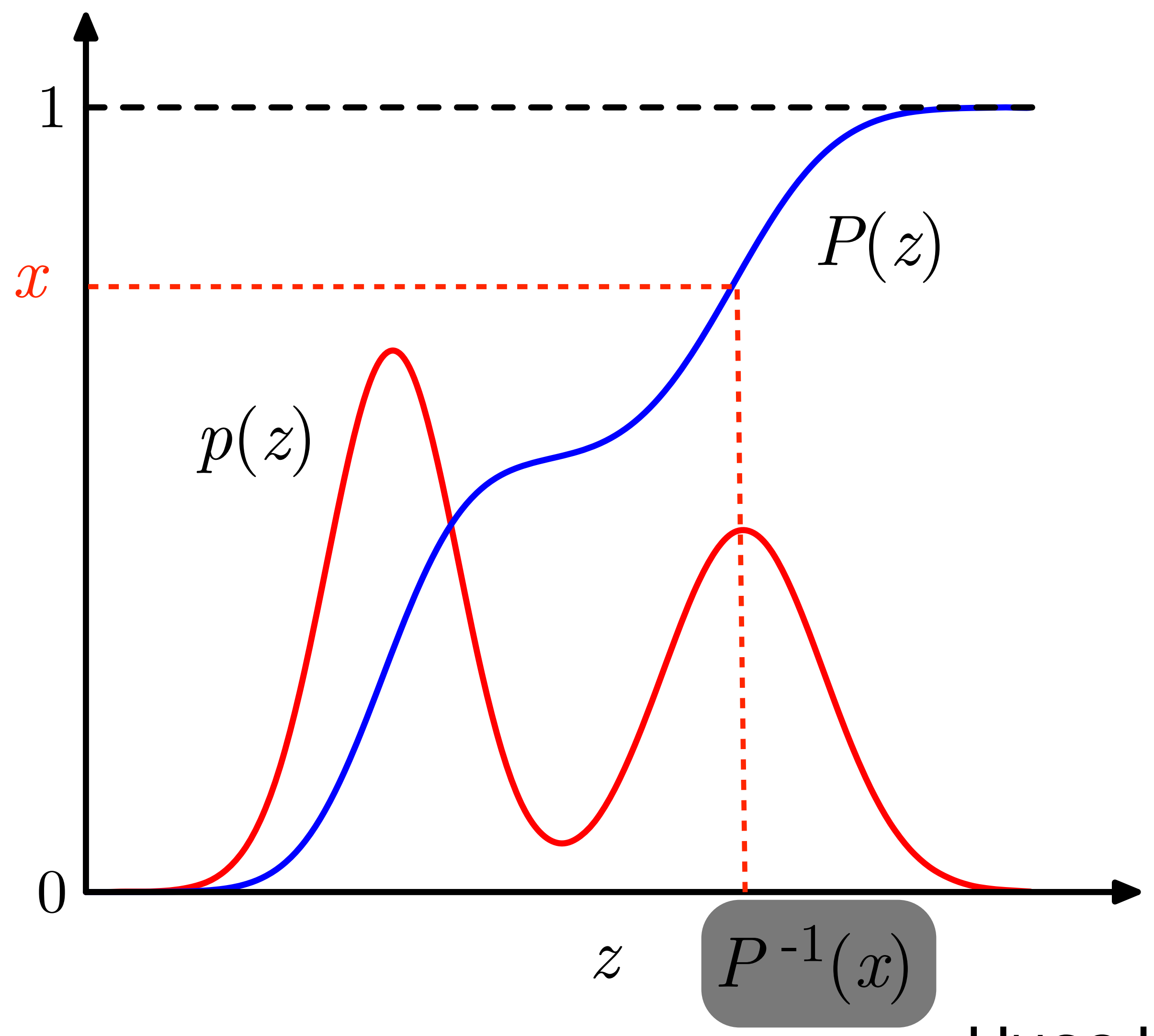
- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **continue** scalaire



ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire continue scalaire

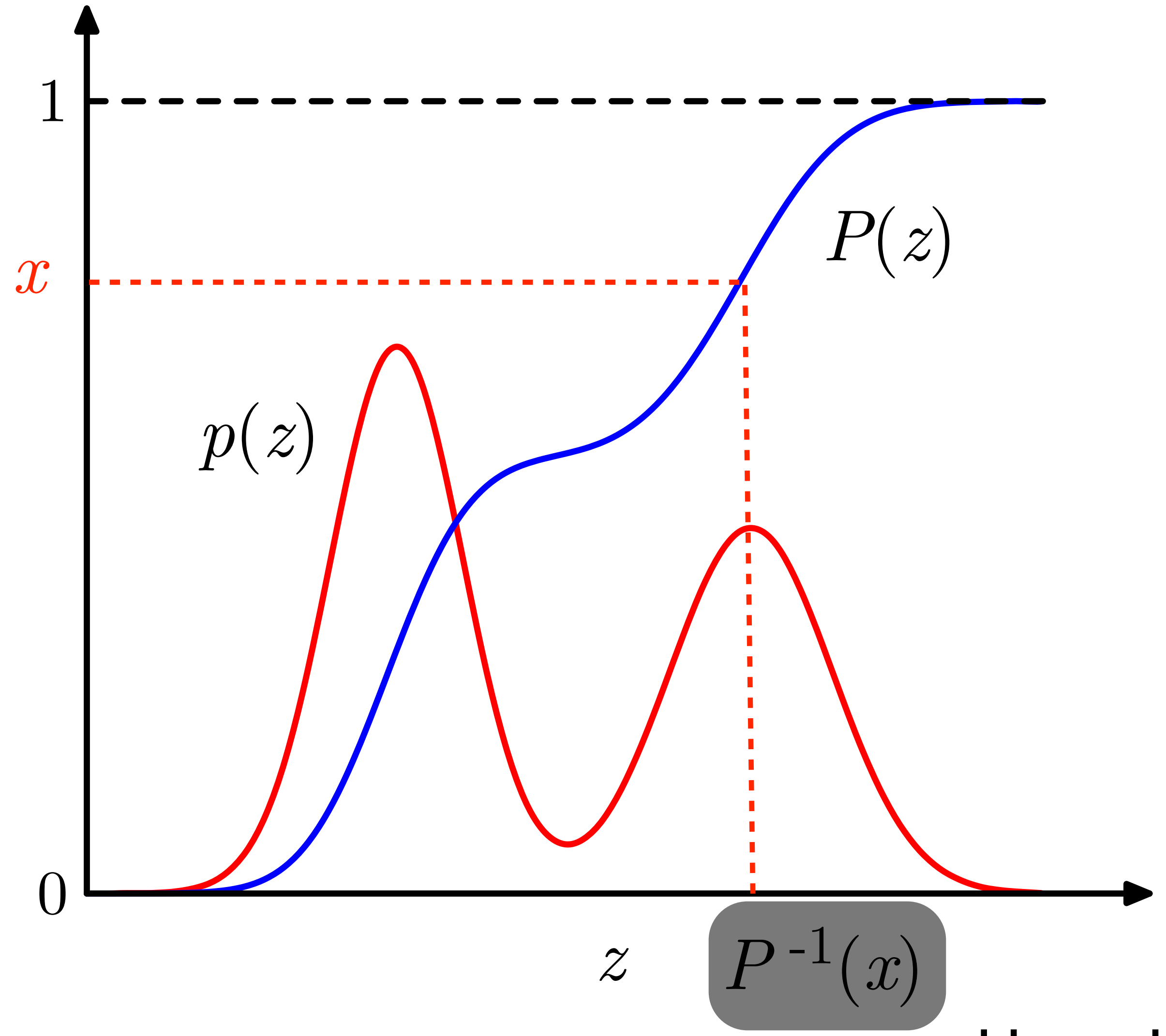
- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **continue** scalaire



ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire continue scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **continue** scalaire



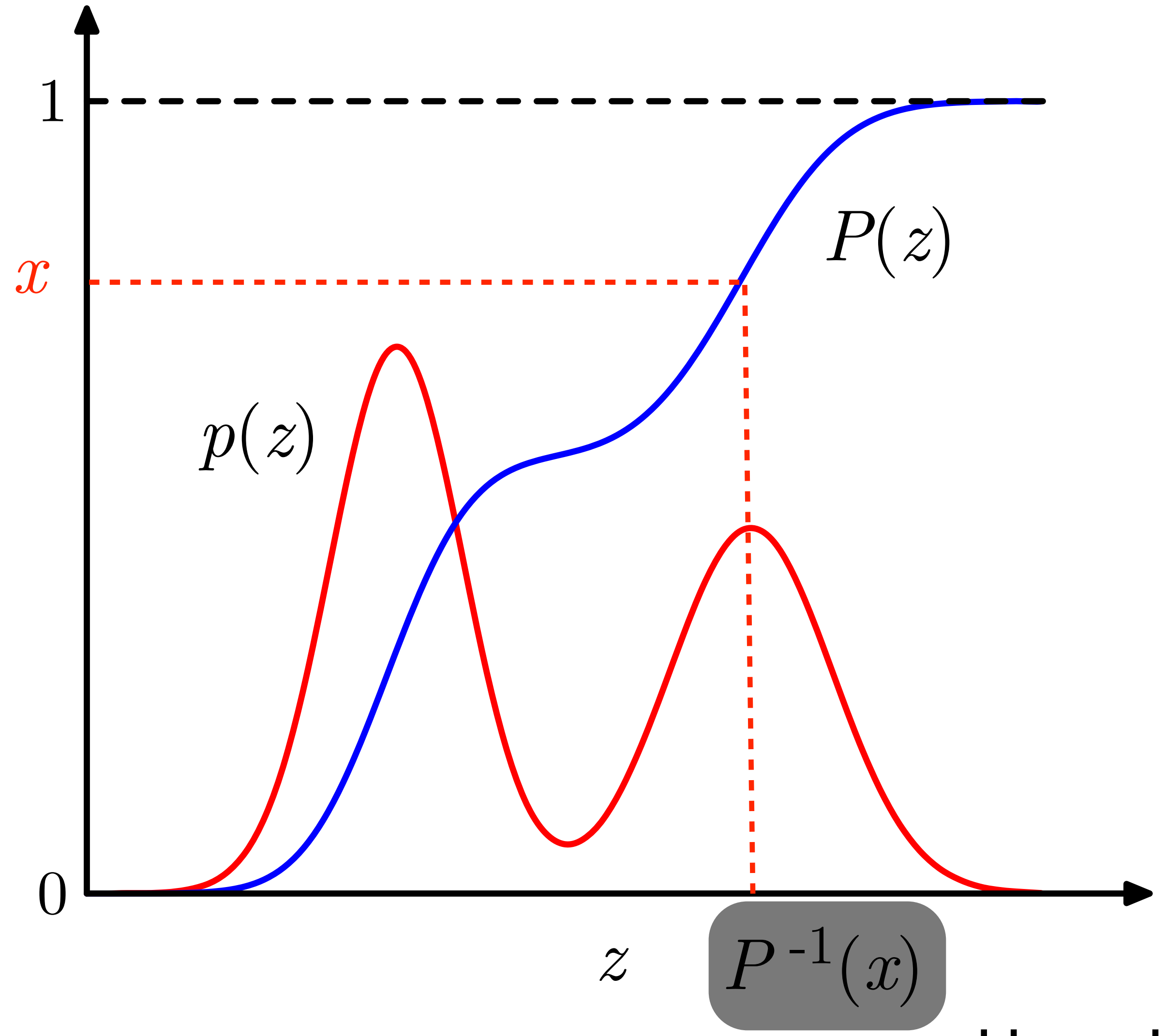
$$p(P^{-1}(X) \leq z)$$

HUGO LAROCHELLE

ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire continue scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **continue** scalaire

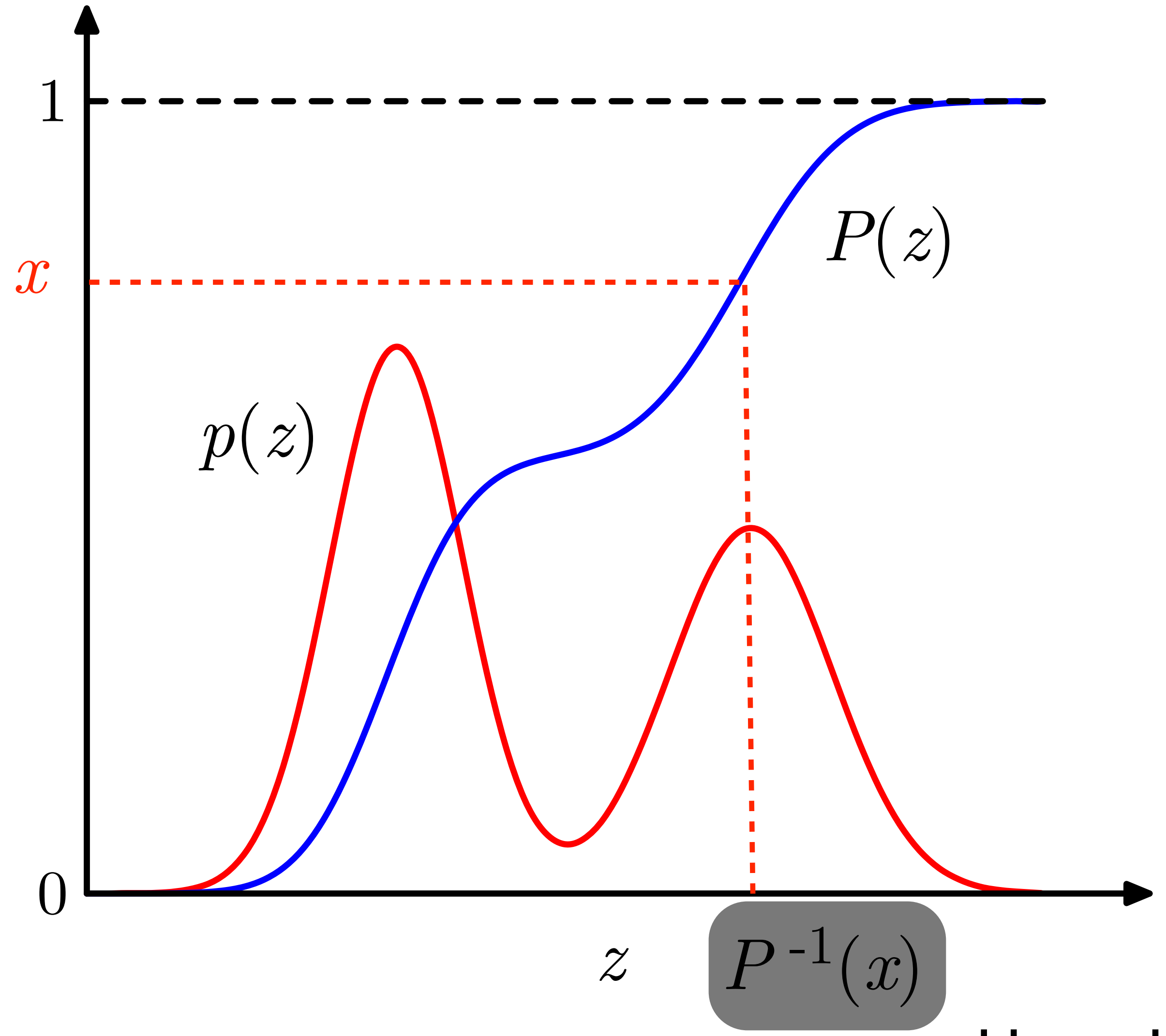


$$p(P^{-1}(X) \leq z) \ = p(X \leq P(z))$$

ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire continue scalaire

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **continue** scalaire



$$p(P^{-1}(X) \leq z) \ = p(X \leq P(z)) = P(z)$$

puisque X est uniforme dans $(0,1)$

Apprentissage automatique

Méthodes d’échantillonnage - vecteur aléatoire gaussien

ÉCHANTILLONNAGE DE BASE

Sujets: méthodes de base

RAPPEL

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire
  - d'une variable aléatoire **continue** scalaire
  - d'une variable aléatoire **gaussienne vectorielle**

- On va supposer qu'on a accès à un générateur pour une variable aléatoire uniformément distribuée entre 0 et 1
  - les méthodes de base vont utiliser les nombres aléatoires qu'il génère

ÉCHANTILLONNAGE DE BASE

Sujets: méthodes de base

RAPPEL

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire
  - d'une variable aléatoire **continue** scalaire
  - d'une variable aléatoire **gaussienne vectorielle**

- On va supposer qu'on a accès à un générateur pour une variable aléatoire uniformément distribuée entre 0 et 1
  - les méthodes de base vont utiliser les nombres aléatoires qu'il génère

ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire gaussienne vectorielle

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **gaussienne vectorielle**

- Variable : Z gaussienne, fonction de densité $\mathcal{N}(\mathbf{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$

- Algorithme

1. générer vecteur gaussien $\mathbf{x}$ de moyenne $\mathbf{0}$ et covariance $\mathbf{I}$
2. calculer $\mathbf{L}$ telle que $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$
3. retourner $\mathbf{L}\mathbf{x} + \boldsymbol{\mu}$

ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire gaussienne vectorielle

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **gaussienne vectorielle**

- Génère de $\mathcal{N}(\mathbf{z}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ puisque
  - transformation linéaire de variables gaussiennes est aussi gaussienne
  - $\mathbb{E}[\mathbf{L}\mathbf{x}+\boldsymbol{\mu}] = \mathbf{L}\mathbb{E}[\mathbf{x}]+\boldsymbol{\mu} = \boldsymbol{\mu}$
  - $\mathrm{cov}[\mathbf{L}\mathbf{x}+\boldsymbol{\mu}] = \mathbf{L}\mathrm{cov}[\mathbf{x}]\mathbf{L}^{\mathrm{T}} = \mathbf{L}\mathbf{L}^{\mathrm{T}} = \boldsymbol{\Sigma}$

ÉCHANTILLONNAGE DE BASE

Sujets: variable aléatoire gaussienne vectorielle

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **gaussienne vectorielle**

- Comment calculer $\mathbf{L}$?
  - décomposition de Cholesky ($\mathbf{L}$ est alors triangulaire inférieure)
  - à partir de la décomposition en valeurs/vecteurs propres de $\mathbf{\Sigma}$:

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{U}^{\mathrm{T}}$$

Apprentissage automatique

Méthodes d’échantillonnage - méthode de rejet

ÉCHANTILLONNAGE DE BASE

Sujets: méthodes de base

RAPPEL

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire
  - d'une variable aléatoire **continue** scalaire
  - d'une variable aléatoire **gaussienne vectorielle**

- On va supposer qu'on a accès à un générateur pour une variable aléatoire uniformément distribuée entre 0 et 1
  - les méthodes de base vont utiliser les nombres aléatoires qu'il génère

ÉCHANTILLONNAGE DE BASE

Sujets: méthodes de base

RAPPEL

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire
  - d'une variable aléatoire **continue** scalaire
  - d'une va

Quoi faire lorsque ces méthodes ne sont pas applicables ?

- On va supposer qu'on a accès à un générateur pour une variable aléatoire uniformément distribuée entre 0 et 1
  - les méthodes de base vont utiliser les nombres aléatoires qu'il génère

Exemple

Sujets: exemple de cas problématique

- Considérons le cas d'une loi

$$p(z) = \frac{1}{Z_p} \widetilde{p}(z)$$

  où calculer Z_p est trop lourd, mais pas $\widetilde{p}(z)$

- Cas spécifiques :
  - loi a posteriori en apprentissage bayésien

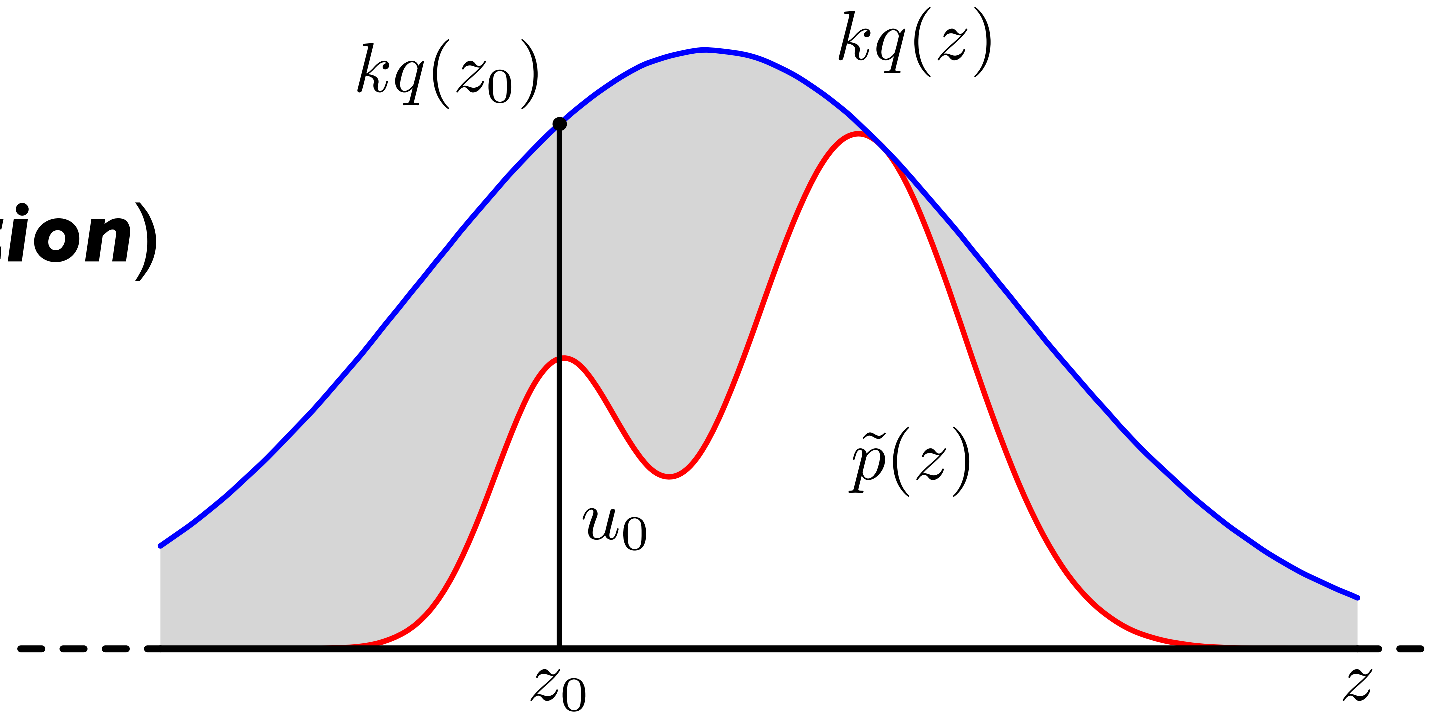
    $$p(\text{«modèle»} \mid \text{«données»}) \propto p(\text{«données»} \mid \text{«modèle»})\ p(\text{«modèle»})$$

  - probabilités $p(\mathbf{z}|\mathbf{x})$ de modèles plus riche qu'un mélange

MÉTHODE DE REJET

Sujets: méthode de rejet, *proposal distribution*

- Supposons qu'on ait une loi $q(z)$ telle que $kq(z) \geqslant \widetilde{p}(z)$ pour une certaine valuer de k
  - $q(z)$ est appelée loi instrumentale (***proposal distribution***)



- **Méthode de rejet**
  1. échantillonner z_0 de $q(z)$
  2. échantillonner u_0 uniformément entre 0 et $kq(z_0)$
  3. si $u_0 < \widetilde{p}(z)$, alors retourner z_0
  4. sinon, revenir à l'étape 1.

MÉTHODE DE REJET

Sujets: méthode de rejet

- La loi $q(z)$ doit être non nulle pour chaque valeur de z telle que $p(z)$ est non nulle
- Si $q(z)$ ne donne pas une borne serrée, cette méthode peut être très inefficace
- Trouver $q(z)$ qui satisfait la condition $kq(z) \geqslant \widetilde{p}(z)$ n'est pas toujours facile
  - il est parfois possible de construire $q(z)$ de façon automatique (voir section 11.1.3)

Apprentissage automatique

Méthodes d’échantillonnage - échantillonnage préférentiel

ÉCHANTILLONNAGE DE BASE

Sujets: méthodes de base

RAPPEL

- Débutons par les méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire
  - d'une variable aléatoire **continue** scalaire
  - d'une va

Quoi faire lorsque ces méthodes ne sont pas applicables ?

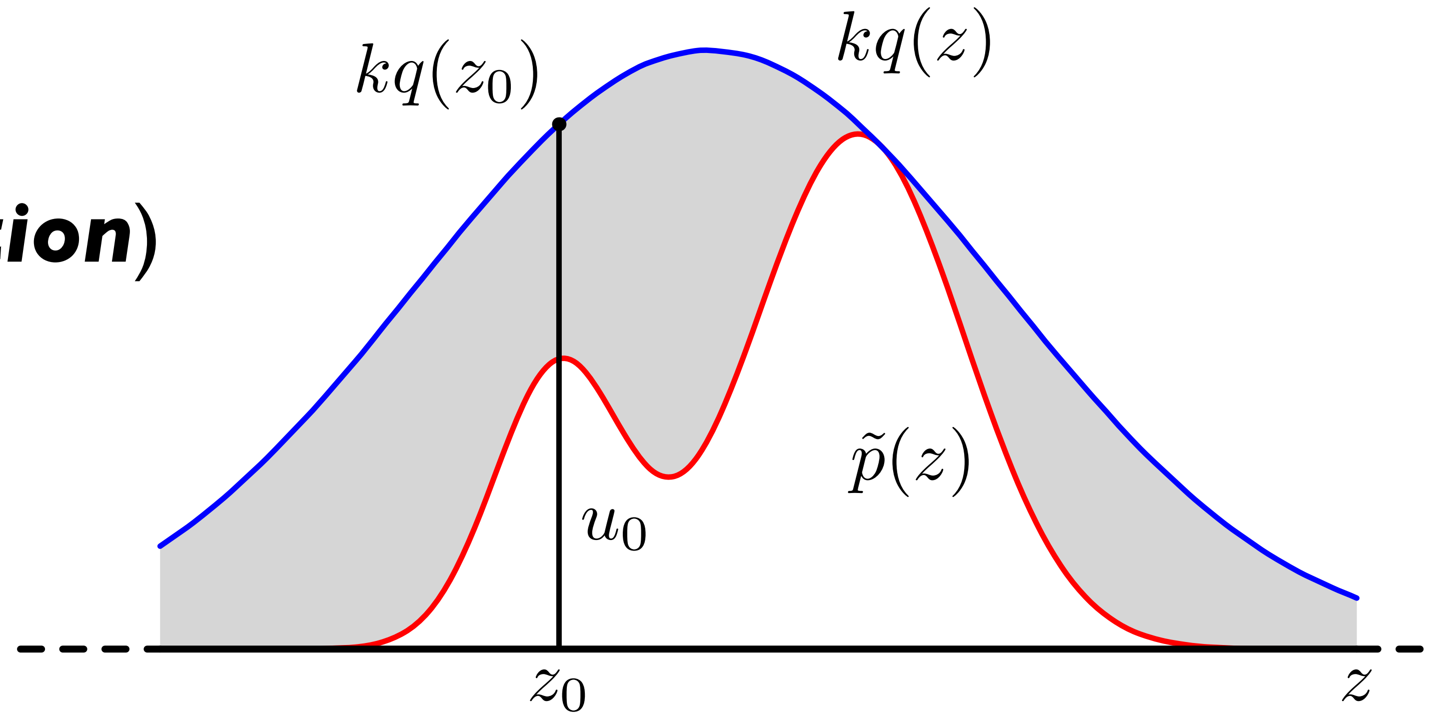
- On va supposer qu'on a accès à un générateur pour une variable aléatoire uniformément distribuée entre 0 et 1
  - les méthodes de base vont utiliser les nombres aléatoires qu'il génère

Méthode de Rejet

Sujets: méthode de rejet, *proposal distribution*

RAPPEL

- Supposons qu'on ait une loi $q(z)$ telle que $kq(z) \geqslant \widetilde{p}(z)$ pour une certaine valuer de k
  - $q(z)$ est appelée loi instrumentale (***proposal distribution***)



- **Méthode de rejet**
  1. échantillonner z_0 de $q(z)$
  2. échantillonner u_0 uniformément entre 0 et $kq(z_0)$
  3. si $u_0 < \widetilde{p}(z)$, alors retourner z_0
  4. sinon, revenir à l'étape 1.

ÉCHANTILLONNAGE

Sujets: estimation d'espérance, estimation Monte Carlo

RAPPEL

- Pourquoi échantillonner d'un modèle ?
  - pour estimer une espérance

$$\mathbb{E}[f] = \int f(\mathbf{z}) p(\mathbf{z}) \, \mathrm{d}\mathbf{z} \;\approx\; \frac{1}{L} \sum_{l=1}^{L} f(\mathbf{z}^{(l)})$$

$\mathbf{z}^{(l)}$ échantillonnés de $p(\mathbf{z})$

- Un tel calcul d'une espérance ainsi est appelé une **estimation Monte Carlo**

ÉCHANTILLONNAGE PRÉFÉRENTIEL

Sujets: échantillonnage préférentiel, *importance weights*

- Pourquoi échantillonner d'un modèle ?
  - pour estimer une espérance

$$\mathbb{E}[f] = \int f(\mathbf{z})p(\mathbf{z})\,\mathrm{d}\mathbf{z}$$

$$= \int f(\mathbf{z})\frac{p(\mathbf{z})}{q(\mathbf{z})}q(\mathbf{z})\,\mathrm{d}\mathbf{z}$$

$$\simeq \frac{1}{L}\sum_{l=1}^{L}\frac{p(\mathbf{z}^{(l)})}{q(\mathbf{z}^{(l)})}f(\mathbf{z}^{(l)}).$$

ÉCHANTILLONNAGE PRÉFÉRENTIEL

Sujets: échantillonnage préférentiel, *importance weights*

- Pourquoi échantillonner d'un modèle ?
  - pour estimer une espérance

$$
\begin{aligned}
\mathbb{E}[f] &= \int f(\mathbf{z})p(\mathbf{z})\,\mathrm{d}\mathbf{z} \\
&= \int f(\mathbf{z})\frac{p(\mathbf{z})}{q(\mathbf{z})}q(\mathbf{z})\,\mathrm{d}\mathbf{z} \\
&\simeq \frac{1}{L}\sum_{l=1}^{L} w_l \; f(\mathbf{z}^{(l)}).
\end{aligned}
$$

Importance weights :

$$w_l = p(\mathbf{z}^{(l)})/q(\mathbf{z}^{(l)})$$

ÉCHANTILLONNAGE PRÉFÉRENTIEL

Sujets: échantillonnage préférentiel, *importance sampling*

- Supposons qu'on ait une loi $q(\mathbf{z})$ (*proposal distribution*)

- **Échantillonnage préférentiel**

1. échantillonner L fois de $q(\mathbf{z})$, pour obtenir $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(L)}$

2. calculer les L poids $w_1, \ldots, w_L$

$$w_l = p(\mathbf{z}^{(l)}) / q(\mathbf{z}^{(l)})$$

3. retourner $\frac{1}{L} \sum_{l=1}^{L} w_l f(\mathbf{z}^l)$

Échantillonnage Préférentiel

Sujets: échantillonnage préférentiel, *importance sampling*

- Comme pour la méthode de rejet, le succès de cette méthode dépend de la proximité entre $q(\mathbf{z})$ et $p(\mathbf{z})$

- Dans le cas où $p(\mathbf{z}) = \frac{1}{Z_p}\widetilde{p}(\mathbf{z})$, on peut utiliser les poids suivants :

$$w_l = \frac{L\ \widetilde{p}(\mathbf{z}^{(l)})/q(\mathbf{z}^{(l)})}{\sum_m \widetilde{p}(\mathbf{z}^{(m)})/q(\mathbf{z}^{(m)})}$$

Apprentissage automatique

Méthodes d’échantillonnage - Metropolis-Hastings

MÉTHODE DE REJET

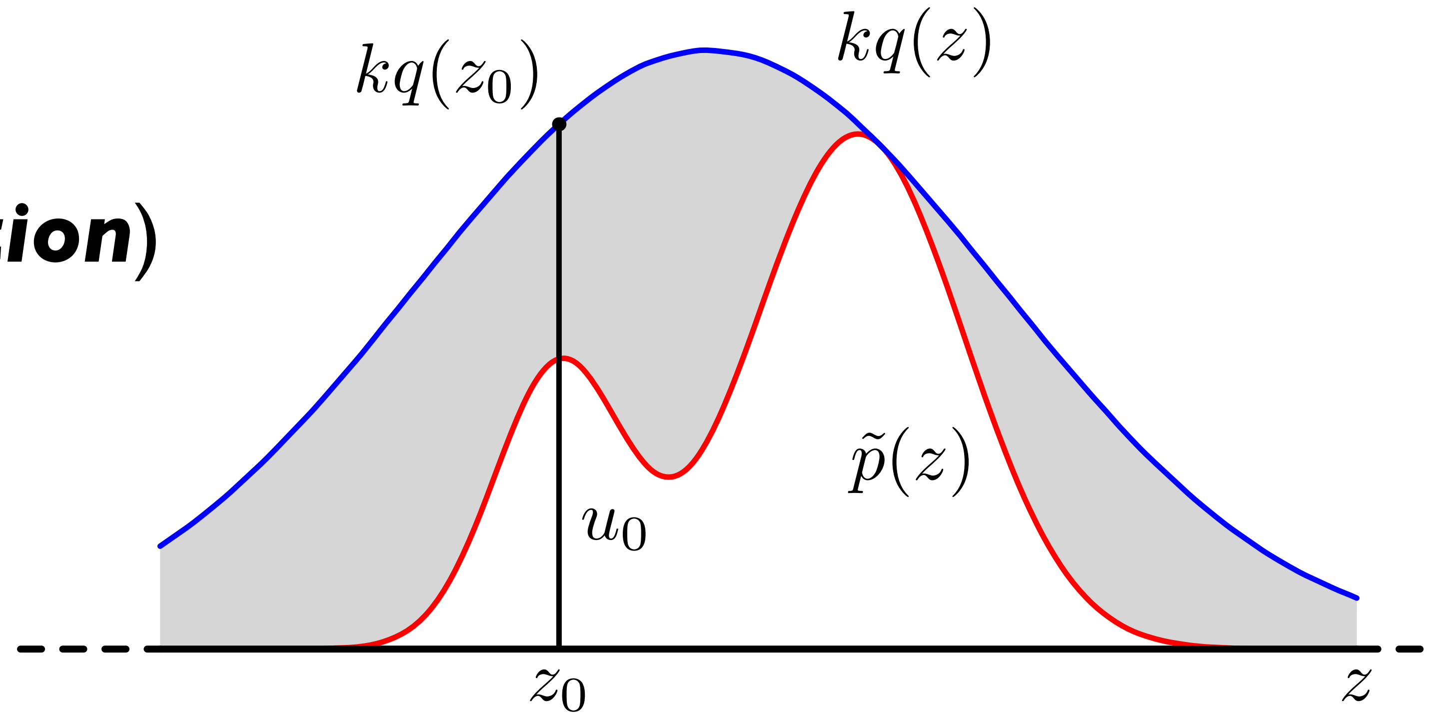
Sujets: méthode de rejet, *proposal distribution*

RAPPEL

- Supposons qu'on ait une loi $q(z)$ telle que $kq(z) \geqslant \widetilde{p}(z)$ pour une certaine valuer de k
  - $q(z)$ est appelée loi instrumentale (***proposal distribution***)

- **Méthode de rejet**

1. échantillonner z_0 de $q(z)$
2. échantillonner u_0 uniformément entre 0 et $kq(z_0)$
3. si $u_0 < \widetilde{p}(z)$, alors retourner z_0
4. sinon, revenir à l'étape 1.



MÉTHODE DE REJET

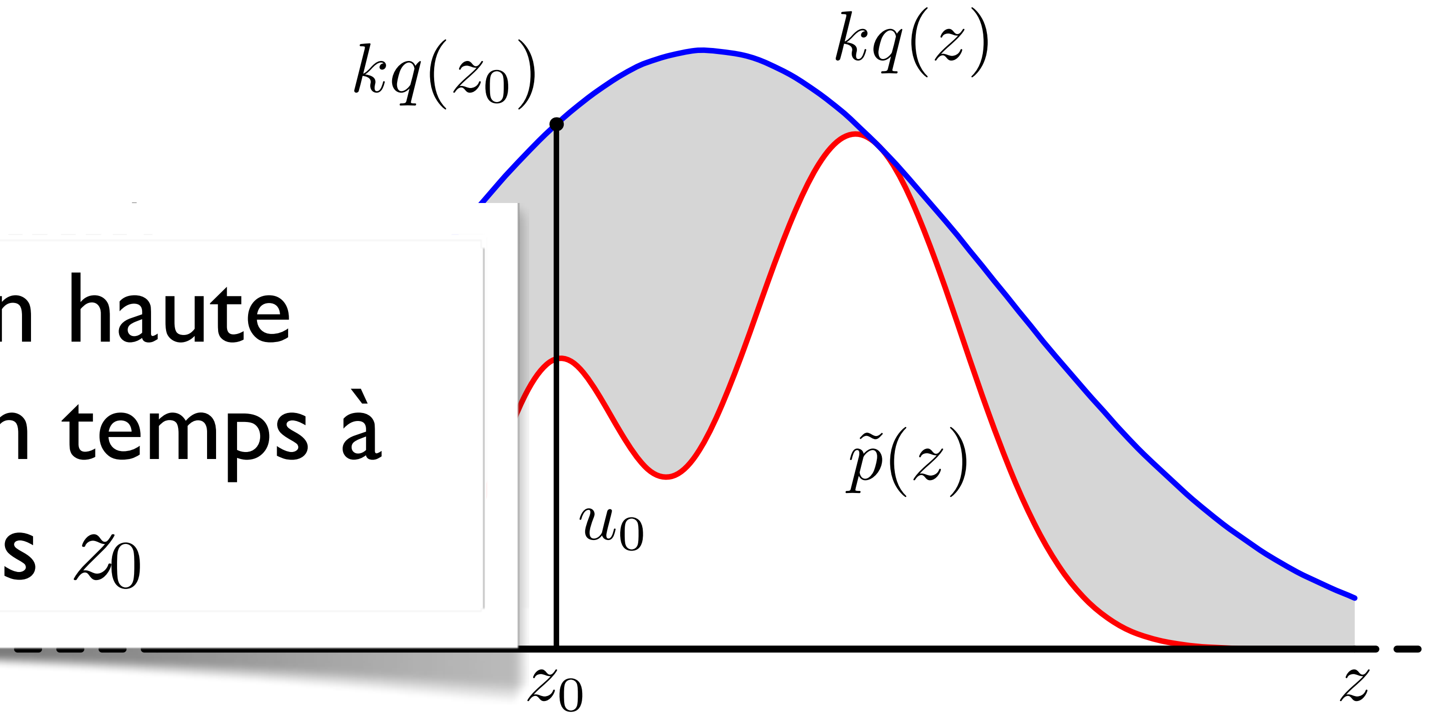
Sujets: méthode de rejet, *proposal distribution*

RAPPEL

- Supposons qu'on ait une loi $q(z)$ telle que $kq(z) \geqslant \widetilde{p}(z)$ pour une certaine valuer de k
  - $q(z)$ est appelée lo

- **Méthode de r**

1. échantillonner z_0 d



Fonctionne souvent mal en haute dimension, puisque passe son temps à rejeter les propositions z_0

2. échantillonner u_0 uniformément entre 0 et $kq(z_0)$
3. si $u_0 < \widetilde{p}(z)$, alors retourner z_0
4. sinon, revenir à l'étape 1.

MONTE CARLO PAR CHAÎNES DE MARKOV

Sujets: MCMC

- On aimerait que l'algorithme tire profit de ses «bons coups»
  - lorsqu'il trouve un bon échantillon z_0, on pourrait l'utiliser pour tirer notre prochain échantillon (p. ex. un échantillon «proche»)
- Spécifiquement, on pourrait tirer les $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(L)}$ en chaîne

$$\mathbf{z}^{(1)} \rightarrow \mathbf{z}^{(2)} \rightarrow \mathbf{z}^{(3)} \rightarrow \cdots \rightarrow \mathbf{z}^{(L)}$$

  - la séquence $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(L)}$ est donc définie par une chaîne de Markov
- C'est ce qu'on appelle une méthode **Monte Carlo par Chaînes de Markov** (MCMC)

METROPOLIS-HASTINGS

Sujets: Metropolis-Hastings, probabilité d'acceptation

- Supposons qu'on ait une loi $q(\mathbf{z}^{(l)}|\mathbf{z}^{(l-1)})$ dernier échantillon généré

- **Metropolis-Hastings**

1. échantillonner un proposition $\mathbf{z}^{*}$ de $q(\mathbf{z}^{(l)}|\mathbf{z}^{(l-1)})$

2. avec **probabilité d'acceptation**

$$A(\mathbf{z}^*, \mathbf{z}^{(l-1)}) = \min\left(1, \frac{\widetilde{p}(\mathbf{z}^*)q(\mathbf{z}^{(l-1)}|\mathbf{z}^*)}{\widetilde{p}(\mathbf{z}^{(l-1)})q(\mathbf{z}^*|\mathbf{z}^{(l-1)})}\right)$$

assigner $\mathbf{z}^{(l)} \leftarrow \mathbf{z}^{*}$, sinon $\mathbf{z}^{(l)} \leftarrow \mathbf{z}^{(l-1)}$

3. retourner $\mathbf{z}^{(l)}$

Metropolis-Hastings

Sujets: Metropolis-Hastings, probabilité d'acceptation

- Supposons qu'on ait une loi $q(\mathbf{z}^{(l)} | \mathbf{z}^{(l\text{-}1)})$ dernier échantillon généré

- **Metropolis-Hastings**

1. échantillonner un proposition $\mathbf{z}^*$ de $q(\mathbf{z}^{(l)} | \mathbf{z}^{(l\text{-}1)})$

2. avec **probabilité d'acceptation**

$$A(\mathbf{z}^*, \mathbf{z}^{(l-1)}) = \min \left(1, \frac{\widetilde{p}(\mathbf{z}^*)}{\widetilde{p}(\mathbf{z}^{(l-1)})} \right)$$

Se simplifie s'il y a symétrie :
$q(\mathbf{z}^{(l)} | \mathbf{z}^{(l\text{-}1)}) = q(\mathbf{z}^{(l\text{-}1)} | \mathbf{z}^{(l)})$

assigner $\mathbf{z}^{(l)} \leftarrow \mathbf{z}^*$, sinon $\mathbf{z}^{(l)} \leftarrow \mathbf{z}^{(l\text{-}1)}$

3. retourner $\mathbf{z}^{(l)}$

Metropolis-Hastings

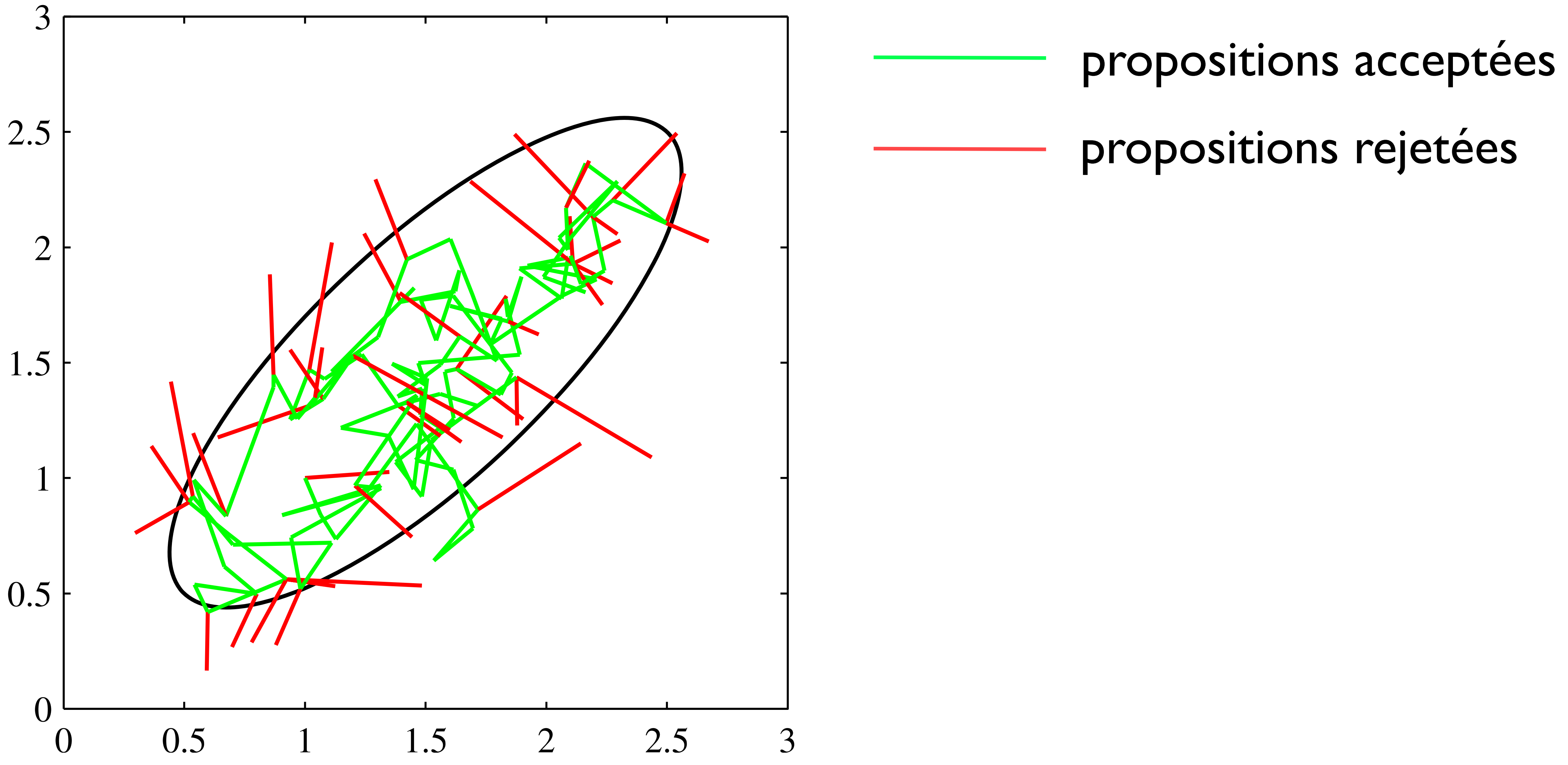
Sujets: Metropolis-Hastings, probabilité d'acceptation

- Le choix de la loi de proposition $q(\mathbf{z}^{(l)}|\mathbf{z}^{(l-1)})$ est important
  - une condition suffisante pour que l'échantillonnage soit valide est que la probabilité de proposer toute valeur légale de $\mathbf{z}$ soit non nulle sous $q(\mathbf{z}^{(l)}|\mathbf{z}^{(l-1)})$
- Un choix courant de $q(\mathbf{z}^{(l)}|\mathbf{z}^{(l-1)})$ est de prendre une gaussienne centrée en $\mathbf{z}^{(l-1)}$
  - une petite covariance impliquera des déplacements lents
  - une grande covariance impliquera des risques de ne pas accepter souvent

METROPOLIS-HASTINGS

Sujets: Metropolis-Hastings

- Exemple :



Apprentissage automatique

Méthodes d’échantillonnage - échantillonnage de Gibbs

Monte Carlo par Chaînes de Markov

Sujets: MCMC

RAPPEL

- On aimerait que l'algorithme tire profit de ses «bons coups»
    - lorsqu'il trouve un bon échantillon z_0, on pourrait l'utiliser pour tirer notre prochain échantillon (p. ex. un échantillon «proche»)
- Spécifiquement, on pourrait tirer les $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(L)}$ en chaîne

$$\mathbf{z}^{(1)} \rightarrow \mathbf{z}^{(2)} \rightarrow \mathbf{z}^{(3)} \rightarrow \ldots \rightarrow \mathbf{z}^{(L)}$$

    - la séquence $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(L)}$ est donc définie par une chaîne de Markov
- C'est ce qu'on appelle une méthode **Monte Carlo par Chaînes de Markov** (MCMC)

Metropolis-Hastings

Sujets: Metropolis-Hastings, probabilité d'acceptation

RAPPEL

- Supposons qu'on ait une loi $q(\mathbf{z}^{(l)}|\mathbf{z}^{(l-1)})$ dernier échantillon généré

- **Metropolis-Hastings**

1. échantillonner un proposition $\mathbf{z}^*$ de $q(\mathbf{z}^{(l)}|\mathbf{z}^{(l-1)})$

2. avec **probabilité d'acceptation**

$$A(\mathbf{z}^*, \mathbf{z}^{(l-1)}) = \min\left(1, \frac{\widetilde{p}(\mathbf{z}^*)q(\mathbf{z}^{(l-1)}|\mathbf{z}^*)}{\widetilde{p}(\mathbf{z}^{(l-1)})q(\mathbf{z}^*|\mathbf{z}^{(l-1)})}\right)$$

assigner $\mathbf{z}^{(l)} \leftarrow \mathbf{z}^*$, sinon $\mathbf{z}^{(l)} \leftarrow \mathbf{z}^{(l-1)}$

3. retourner $\mathbf{z}^{(l)}$

Metropolis-Hastings

Sujets: Metropolis-Hastings, probabilité d'acceptation

RAPPEL

- Supposons qu'on ait une loi $q(\mathbf{z}^{(l)}\,|\mathbf{z}^{(l\text{-}1)})$ dernier échantillon généré

- **Metropolis-Hastings**

1. échantillonner un proposition $\mathbf{z}^*$ de $q(\mathbf{z}^{(l)}\,|\mathbf{z}^{(l\text{-}1)})$

2. avec **probabilité d'acc**

Peut-on éviter d'avoir à fournir une loi $q(\mathbf{z}_l\,|\mathbf{z}_{l\text{-}1})$?

$$A(\mathbf{z}^*, \mathbf{z}^{(l-1)}) = \min\left(1, \frac{p(\mathbf{z}^*)q(\mathbf{z}^{(l-1)}|\mathbf{z}^*)}{\widetilde{p}(\mathbf{z}^{(l-1)})q(\mathbf{z}^*|\mathbf{z}^{(l-1)})}\right)$$

assigner $\mathbf{z}^{(l)} \leftarrow \mathbf{z}^*$, sinon $\mathbf{z}^{(l)} \leftarrow \mathbf{z}^{(l\text{-}1)}$

3. retourner $\mathbf{z}^{(l)}$

ÉCHANTILLONNAGE DE GIBBS

Sujets: Metropolis-Hastings, probabilité d'acceptation

- **Échantillonnage de Gibbs**

1. assigner $\mathbf{z}^{(l)} \leftarrow \mathbf{z}^{(l\text{-}1)}$
2. choisir k aléatoirement parmi $\{1, \ldots, D\}$
3. remplacer $z_k^{(l)}$ par un échantillon de $p(z_k^{(l)} | z_1^{(l)}, \ldots, z_{k-1}^{(l)}, z_{k+1}^{(l)}, \ldots, z_D^{(l)})$
4. retourner $\mathbf{z}^{(l)}$

Échantillonnage de Gibbs

Sujets: Metropolis-Hastings, probabilité d'acceptation

- **Échantillonnage de Gibbs**

1. assigner $\mathbf{z}^{(l)} \leftarrow \mathbf{z}^{(l\text{-}1)}$
2. choisir k aléatoirement parmi $\{1, \ldots, D\}$
3. remplacer $z_k^{(l)}$ par un échantillon de $p(z_k^{(l)} | \underbrace{z_1^{(l)}, \ldots, z_{k-1}^{(l)}, z_{k+1}^{(l)}, \ldots, z_D^{(l)}}_{\mathbf{z}_{\backslash k}^{(l)}})$
4. retourner $\mathbf{z}^{(l)}$

$\mathbf{z}_{\backslash k}^{(l)}$: tout le vecteur $\mathbf{z}^{(l)}$ sans la k^{e} composante

ÉCHANTILLONNAGE DE GIBBS

Sujets: Metropolis-Hastings, probabilité d'acceptation

- Équivaut au Metropolis-Hastings, mais dont la loi instrumentale est telle qu'on accepte toujours

$$\frac{p(\mathbf{z}^*)q(\mathbf{z}^{(l-1)}|\mathbf{z}^*)}{p(\mathbf{z}^{(l-1)})q(\mathbf{z}^*|\mathbf{z}^{(l-1)})} = \frac{p(z_k^*|\mathbf{z}^*_{\setminus k})p(\mathbf{z}^*_{\setminus k})p(z_k^{(l-1)}|\mathbf{z}^*_{\setminus k})}{p(z_k^{(l-1)}|\mathbf{z}^{(l-1)}_{\setminus k})p(\mathbf{z}^{(l-1)}_{\setminus k})p(z_k^*|\mathbf{z}^{(l-1)}_{\setminus k})}$$

ÉCHANTILLONNAGE DE GIBBS

Sujets: Metropolis-Hastings, probabilité d'acceptation

- Équivaut au Metropolis-Hastings, mais dont la loi instrumentale est telle qu'on accepte toujours

$$\mathbf{z}_{\backslash k}^{(l-1)} = \mathbf{z}_{\backslash k}^{*}$$

$$\frac{p(\mathbf{z}^*)q(\mathbf{z}^{(l-1)}|\mathbf{z}^*)}{p(\mathbf{z}^{(l-1)})q(\mathbf{z}^*|\mathbf{z}^{(l-1)})} = \frac{p(z_k^*|\mathbf{z}_{\backslash k}^*)p(\mathbf{z}_{\backslash k}^*)p(z_k^{(l-1)}|\mathbf{z}_{\backslash k}^*)}{p(z_k^{(l-1)}|\mathbf{z}_{\backslash k}^{(l-1)})p(\mathbf{z}_{\backslash k}^{(l-1)})p(z_k^*|\mathbf{z}_{\backslash k}^{(l-1)})}$$

ÉCHANTILLONNAGE DE GIBBS

Sujets: Metropolis-Hastings, probabilité d'acceptation

- Équivaut au Metropolis-Hastings, mais dont la loi instrumentale est telle qu'on accepte toujours

$$\mathbf{z}_{\setminus k}^{(l-1)} = \mathbf{z}_{\setminus k}^{*}$$

$$\frac{p(\mathbf{z}^*)q(\mathbf{z}^{(l-1)}|\mathbf{z}^*)}{p(\mathbf{z}^{(l-1)})q(\mathbf{z}^*|\mathbf{z}^{(l-1)})} = \frac{p(z_k^*|\mathbf{z}_{\setminus k}^*)p(\mathbf{z}_{\setminus k}^*)p(z_k^{(l-1)}|\mathbf{z}_{\setminus k}^*)}{p(z_k^{(l-1)}|\ \mathbf{z}_{\setminus k}^*\ )p(\ \mathbf{z}_{\setminus k}^*\ )p(z_k^*|\ \mathbf{z}_{\setminus k}^*\ )}$$

ÉCHANTILLONNAGE DE GIBBS

Sujets: Metropolis-Hastings, probabilité d'acceptation

- Équivaut au Metropolis-Hastings, mais dont la loi instrumentale est telle qu'on accepte toujours

$$\boxed{\mathbf{z}_{\setminus k}^{(l-1)} = \mathbf{z}_{\setminus k}^{*}}$$

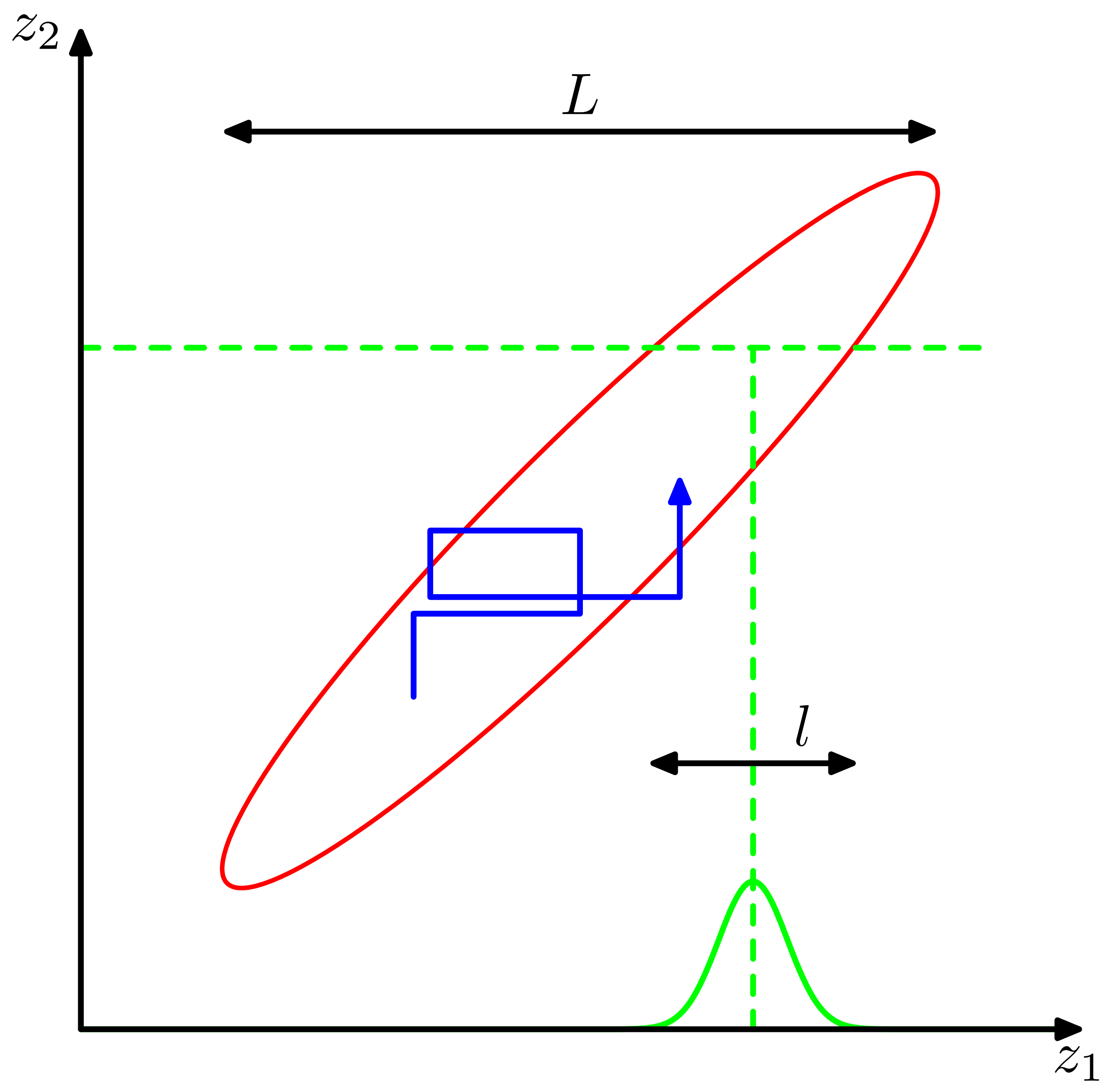
$$\frac{p(\mathbf{z}^*)q(\mathbf{z}^{(l-1)}|\mathbf{z}^*)}{p(\mathbf{z}^{(l-1)})q(\mathbf{z}^*|\mathbf{z}^{(l-1)})} = \frac{p(z_k^*|\mathbf{z}_{\setminus k}^*)p(\mathbf{z}_{\setminus k}^*)p(z_k^{(l-1)}|\mathbf{z}_{\setminus k}^*)}{p(z_k^{(l-1)}|\ \mathbf{z}_{\setminus k}^*\ )p(\ \mathbf{z}_{\setminus k}^*\ )p(z_k^*|\ \mathbf{z}_{\setminus k}^*\ )}$$

$$= 1$$

ÉCHANTILLONNAGE DE GIBBS

Sujets: Metropolis-Hastings, probabilité d'acceptation

- Exemple :



ÉCHANTILLONNAGE DE GIBBS

Sujets: version cyclique

- **Échantillonnage de Gibbs cyclique**

  1. assigner $\mathbf{z}^{(l)} \leftarrow \mathbf{z}^{(l\text{-}1)}$

  2. pour k de 1 à D

     - remplacer $z_k^{(l)}$ par un échantillon de $p(z_k^{(l)} | z_1^{(l)}, \ldots, z_{k-1}^{(l)}, z_{k+1}^{(l)}, \ldots, z_D^{(l)})$

  3. retourner $\mathbf{z}^{(l)}$

- Permet de réduire la corrélation entre les $\mathbf{z}^{(l)}$

- L'ordre de remplacement des $z_k^{(l)}$ n'est pas important

Apprentissage automatique

Méthodes d’échantillonnage - résumé

ÉCHANTILLONNAGE

Sujets: estimation d'espérance, estimation Monte Carlo

RAPPEL

- Pourquoi échantillonner d'un modèle ?
  - pour estimer une espérance

$$\mathbb{E}[f] = \int f(\mathbf{z})p(\mathbf{z})\,\mathrm{d}\mathbf{z} \;\approx\; \frac{1}{L}\sum_{l=1}^{L} f(\mathbf{z}^{(l)})$$

$\mathbf{z}^{(l)}$ échantillonnés de $p(\mathbf{z})$

- Un tel calcul d'une espérance ainsi est appelé une **estimation Monte Carlo**

ÉCHANTILLONNAGE DE BASE

Sujets: méthodes de base

- Méthodes de base pour l'échantillonnage
  - d'une variable aléatoire **discrète** scalaire
  - d'une variable aléatoire **continue** scalaire
  - d'une variable aléatoire **gaussienne vectorielle**

- Dans ces cas, on est garanti d'obtenir des échantillons de qualité, en un temps bien déterminé

MÉTHODE DE REJET

Sujets: méthode de rejet

- Si les méthodes de base ne sont pas applicables à $p(\mathbf{z})$:

Méthode de rejet :

- doit fournir une loi instrumentale (*proposal distribution*) $q(\mathbf{z})$
- si la loi $q(\mathbf{z})$ n'est pas proche de $p(\mathbf{z})$, on peut attendre longtemps avant d'avoir un échantillon (particulièrement en haute dimension)

ÉCHANTILLONNAGE PRÉFÉRENTIEL

Sujets: échantillonnage préférentiel

- Si les méthodes de base ne sont pas applicables à $p(\mathbf{z})$:

Échantillonnage préférentiel

- doit fournir une loi instrumentale (*proposal distribution*) $q(\mathbf{z})$
- meilleur que la méthode de rejet pour l'estimation Monte Carlo
- permet seulement d'approximer une somme (ne donne pas des échantillons de $p(\mathbf{z})$)
- si la loi $q(\mathbf{z})$ n'est pas proche de $p(\mathbf{z})$, l'approximation peut être très mauvaise (particulièrement en haute dimension)

Metropolis-Hastings

Sujets: Metropolis-Hastings

- Si les méthodes de base ne sont pas applicables à $p(\mathbf{z})$:

Métropolis-Hastings

- méthode MCMC
- nécessite seulement une loi instrumentale conditionnelle $q(\mathbf{z}^{(l)} \,|\mathbf{z}^{(l-1)})$
- donne une séquence d'échantillons qui sont corrélés (ne peut pas les traiter comme des échantillons indépendants)

ÉCHANTILLONNAGE DE GIBBS

Sujets: Échantillonnage de Gibbs

- Si les méthodes de base ne sont pas applicables à $p(\mathbf{z})$:

Échantillonnage de Gibbs

- méthode MCMC
- ne nécessite pas de loi instrumentale
- nécessite que les conditionnelles $p(z_k^{(l)}|z_1^{(l)}, \dots, z_{k-1}^{(l)}, z_{k+1}^{(l)}, \dots, z_D^{(l)})$ soient calculables
- donne une séquence d'échantillons qui sont corrélés (ne peut pas les traiter comme des échantillons indépendants)

Metropolis-Hastings

Sujets: Metropolis-Hastings

- Toutes les méthodes MCMC génèrent des échantillons corrélés
  - pour réduire la corrélation, on collecte seulement tous les M échantillons (p. ex. $M{=}100$)
  - on ignore également les premiers échantillons générés (période de *burn-in*)

- Pour en savoir plus sur les propriétés que doivent satisfaire les méthodes MCMC : voir section 11.2.1