Stochastik–Praktikum
Varianzreduktions Methoden bei Monte Carlo

Dirk Becherer

Humboldt-Universität zu Berlin

14. Januar 2015

Übersicht

1. Konfidenzintervalle bei Monte Carlo
2. Antithetic Variates
3. Control Variates
4. Importance Sampling

Übersicht

1 Konfidenzintervalle bei Monte Carlo

2 Antithetic Variates

3 Control Variates

4 Importance Sampling

Genauigkeit von Monte Carlo Schätzern

- Ziel: Approximative Bestimmung von Erwartungswert $E[X]$ für intergrierbare Zufallsvariable $X = h(Y)$ durch Monte Carlo Simulation, basierend auf i.i.d. Stichprobe $X_i = h(Y_i)$, $i = 1, \ldots, n$ vom Umfang n.
- **Gewöhnlicher Monte Carlo Schätzer** für $E[X]$:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} h(Y_i)$$

- Wissen $\bar{X}_n \to E[X]$ f.s. für $n \to \infty$ (Starkes GdgZ)
- Wie gut ist die **zufällige** Approximation $\bar{X}_n$?

Konfidenzintervalle

- Realisierte Fehler $E[X] - \bar{X}_n(\omega)$: unbekannt, wenn MC nötig.
- Ein $(1-\alpha)$- **Konfidenzintervall** ist ein Bereich $[a, b]$ als Funktion der Stichprobe $X_{1:n} := (X_1, \ldots, X_m)$, welcher (vor Realisierung!) den unbekannten Parameter $m_X := E[X]$ mit vorgegebener Wahrscheinlichkeit (=Vertrauensniveau) $1-\alpha$ umfasst:

$$P[a(X_{1:n}) \leq m_X \leq b(X_{1:n})] \geq 1-\alpha$$

- Anwendungsbezogene Wahl des Konfidenz**niveaus**, z.B. $1-\alpha = 95\%, 99\%$ oder $99,99\%$ für Irrtumswahrscheinlichkeit (Irrtumsniveau) $\alpha = 5\%,\ 1\%$ bzw. 0.01%.
- Bestimmung von (approximativen) Konfidenzintervallen ?

Konfidenzintervalle via ZGWS

- Zentraler Grenzwertsatz : $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ ist approximativ $\mathcal{N}(m_X, \sigma_X^2/n)$-verteilt.
- $\rightsquigarrow$ Grenzen des (kleinsten, approx.) Konfidenzintervalles $[a, b]$ aus Quantilen $q(x)$ der $\mathcal{N}(0, 1)$-Normalverteilung:

$$a(X_1, \ldots, X_n) = \bar{X}_n - q(1 - \frac{\alpha}{2})\frac{s_n}{\sqrt{n}}$$

$$b(X_1, \ldots, X_n) = \bar{X}_n + q(1 - \frac{\alpha}{2})\frac{s_n}{\sqrt{n}}$$

- Dabei wird die i.a. unbekannte Varianz σ_X^2 erwartungstreu geschätzt durch $s_n^2 := \frac{1}{n-1}\sum_1^n (X_i - \bar{X}_n)^2$. (Matlab: std)
- **Konfidenzintervalllänge** skaliert mit $s_n/\sqrt{n} \approx \sigma_N/\sqrt{n}$
- Quantil Bsp: $q(1 - \alpha/2) = q(0.975) \approx 1.96$ für $\alpha = 5\%$

Übersicht

1 Konfidenzintervalle bei Monte Carlo

2 Antithetic Variates

3 Control Variates

4 Importance Sampling

Antithetic Variates

Idee: Erzeuge i.i.d. Folge von Paaren von Zufallsvariablen

$$(X_i, \tilde{X}_i) = (h(Y_i), h(\tilde{Y}_i))\,, \quad i = 1, 2, \ldots$$

so dass

- die antithetische Variablen $\tilde{Y}_i$ bzw. $\tilde{X}_i$ sich einfach (deterministisch) aus Y_i bzw. X_i erzeugen lassen,
- die **Paare** $(X_i, \tilde{X}_i) = (h(Y_i), h(\tilde{Y}_i))$ **i.i.d.** sind,
- X_i und $\tilde{X}_i$ (bzw. Y_i und $\tilde{Y}_i$) diegleiche Verteilung haben, aber i.a. **nicht** unabhängig sind.
- Dann hat $X_i^{\mathrm{av}} := \frac{1}{2}(X_i + \widetilde{X}_i)$ dengleichen Erwartungswert: $E[X^{\mathrm{av}}] = E[X] = E[h(Y)]$
- Der **Antithetische Monte Carlo** Schätzer dafür ist

$$\bar{X}_n^{\mathrm{av}} = \frac{1}{n}\sum_{i=1}^{n} X_i^{\mathrm{av}} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}(h(Y_i) + h(\widetilde{Y}_i))$$

Varianzreduktion durch Antithetic Variates

Beispiele: $\tilde{Y}_i := 1 - Y_i$ für $Y_i \sim U[0,1]$ iid,
oder $\tilde{Y}_i := -Y_i$ für $Y_i \sim \mathcal{N}(0,1)$ iid.
Vorteile

- Die Paare zu erzeugen erfordert kaum extra Aufwand
- **Varianzreduktion** falls $Var(X^{\mathrm{av}}) = Var\big((h(Y) + h(\widetilde{Y}))/2\big)$ kleiner als $Var(X) = Var(h(Y))$ ist

$$
\begin{aligned}
Var(X^{\mathrm{av}}) &= \frac{1}{2}\Big(Var(h(Y)) + Cov(h(Y), h(\widetilde{Y})\Big) \\
&= \frac{1}{2}\Big(Var(X) + Cov(X, \widetilde{X})\Big)
\end{aligned}
$$

Varianzreduktion

Beispiele: $\tilde{Y}_i := 1 - Y_i$ für $Y_i \sim U[0,1]$ iid,
oder $\tilde{Y}_i := -Y_i$ für $Y_i \sim \mathcal{N}(0,1)$ iid.

Satz

Ist h monoton (wachsend oder fallend) und $\widetilde{Y} = g(Y)$ für g fallend, dann $Cov(h(Y), h(\widetilde{Y})) \leq 0$ und $Var(X^{\mathrm{av}}) \leq \frac{1}{2} Var(X)$

Beweis:
$2Cov(h(Y), -h(g(Y))) =$
$E[(h(Y_1) - h(Y_2)(-h(g(Y_1)) + h(g(Y_2))] \geq 0$

Beispiel für MC Antithetische Variablen

Beispiel: MC Simulation zur (testweisen) Berechnung von $E[X]$ für $X = exp(Z)$ mit $Y = Z \sim \mathcal{N}(0,1)$
Hier bereits bekannt $E[X] = exp(1/2)$

Vergleich der Approximationsgüte vom einfachen MC Schätzer $\bar{X}_n$ basierend auf $X_1, \dots, X_n$ vs. antithetischen MC Schätzer $\bar{X}_n^{\mathrm{av}}$ basierend auf $X_1^{\mathrm{av}}, \dots, X_n^{\mathrm{av}}$ durch

- Längenverhältnis der 95%-Konfidenzintervalle
- also gleich Verhältnis der Standardabweichungen s_n

Matlab: Antithetic Variates Example

```
Variance reduction experiment with Antithetic Variables
for E[exp(Z)] for Z standardnormal
True expectation known =1.64872
```

	log10(n)	MCmean	MCKI	KIwidth	WidthRatioAntiVsPlain
PlainMC :	log10(n)	MCmean	MCKI	KIwidth	
	2	1.181	[0.924,1.438]	0.514	
	3	1.647	[1.512,1.782]	0.271	
	4	1.661	[1.619,1.704]	0.086	
	5	1.648	[1.635,1.662]	0.026	
Antithetic :	log10(n)	MCmean	MCKI	KIwidth	WidthRatioAntiVsPlain
	2	1.484	[1.330,1.638]	0.308	5.995e−01
	3	1.647	[1.574,1.721]	0.147	5.442e−01
	4	1.667	[1.642,1.691]	0.049	5.707e−01
	5	1.650	[1.642,1.657]	0.015	5.640e−01

Multivariate anthithetische Variablen

Multivariater Fall : Sei $X = h(Y) = h(Y^1, \dots, Y^d)$.

Beispiele antithetischer Variablen:

- Für $Y = (Y^1, \dots, Y^d)$ mit $Y^i \sim U[0,1]$ iid. ist antithetische Variable $\widetilde{Y} = (1 - Y^1, \dots, 1 - Y^d)$
- Für $Y = (Y^1, \dots, Y^d)$ mit $Y^i \sim \mathcal{N}(0,1)$ iid. ist antithetische Variable $\widetilde{Y} = -Y = (-Y^1, \dots, -Y^d)$

Analog zum univariaten Fall zeigt auch hier der antithetische MC Schätzer $\bar{X}_n^{\mathrm{av}}$ basierend auf Simulationen von

$$X^{\mathrm{av}} = \frac{1}{2}(h(Y) + h(\widetilde{Y}))$$

reduzierte Varianz im Vergleich zu $\bar{X}_n$, falls h monoton in allen Argumenten ist (in jedem jeweils wachsend oder fallend).

Übersicht

1. Konfidenzintervalle bei Monte Carlo
2. Antithetic Variates
3. Control Variates
4. Importance Sampling

Control Variates

Idee: Eine ähnliche Zufallsvariable X^{cv}, die **Kontrollvariable**, deren Erwartungswert man schon kennt, benutzen, und nur den Erwartungswert der Differenz $X - X^{\text{cv}}$ durch MC Simulation berechnen
$\Rightarrow$ Varianzreduktion falls $X - X^{\text{cv}}$ ‘kleiner‘ als X ist.

Vereinfachter Fall (ohne Regressionsparameter b):

$$\begin{aligned} X' &:= X - (X^{\text{cv}} - E[X^{\text{cv}}]) \\ E[X'] &= E[X] = E[X - X^{\text{cv}}] + E[X^{\text{cv}}] \\ Var[X'] &= Var[X - X^{\text{cv}}] \end{aligned}$$

MC Schätzer für X' bzw. $X - X^{\text{cv}}$ ist CV-MC Schätzer für X

Control Variates

Allgemeiner Fall nutzt zusätzlich Regression mit $b \in \mathbb{R}$

$$X'_b := X - b(X^{\mathrm{cv}} - E[X^{\mathrm{cv}}])$$

$$E[X'_b] = E[X] = E[X - bX^{\mathrm{cv}}] + bE[X^{\mathrm{cv}}]$$

Die Wahl $b^* = Cov(X, X^{\mathrm{cv}})/Var(X^{\mathrm{cv}})$ minimiert $Var(X_b)$

$$Var(X'_{b*}) = Var(X) - (Cov(X, X^{\mathrm{cv}}))^2/Var(X^{\mathrm{cv}}) \leq Var(X)$$

Varianzreduktion umso größer, je stärker X^{cv} korreliert zu X ist. Ergebnis ist dank b invariant unter affinen Transformationen $X_2^{\mathrm{cv}} = A + BX_1^{\mathrm{cv}}$ ($B \neq 0$) der Kontrollvariablen.

Regressionsparameter Schätzen

In der Regel ist $b^* = Cov(X, X^{\mathrm{cv}})/Var(X^{\mathrm{cv}})$ nicht bekannt.

Daher wird b^* praktisch als die Steigung der Least-Squares Regressionsgraden aus den Simulationswerten von X und X^{cv} geschätzt (in Matlab mit 'polyfit', dh. als Quotient $(s_{X,X^{\mathrm{cv}}})^2/(s_{X^{\mathrm{cv}}})^2$ der erwartungstreuen Schätzer $(s_{X,X^{\mathrm{cv}}})^2$, $(s_{X^{\mathrm{cv}}})^2$ für $Cov(X, X^{\mathrm{cv}})$ bzw. $Var(X^{\mathrm{cv}})$).

Im Prinzip könnte die Varianz dabei wg. Schätzfehlern auch größer werden - es funktioniert aber meist gut.

Beispiele versch. Control Variates

Ziel: Berechnung von $E[X]$ für

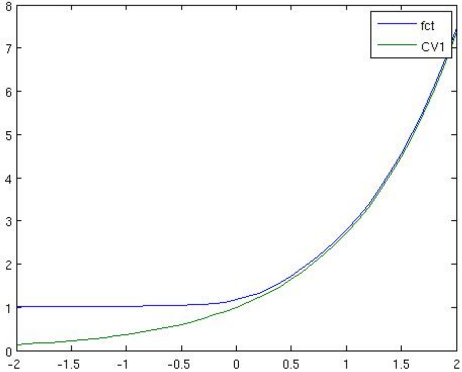
$$X = \exp(\frac{1}{2}(Z + \sqrt{Z^2 + 0.1}))$$

für Z als $\mathcal{N}(0, 1)$-normalverteilt.
Verschiedene mögliche Kontrollvariablen

- $X_1^{\mathrm{cv}} = \exp(Z)$
- $X_2^{\mathrm{cv}} = 5 * X_1^{\mathrm{cv}} + 2 = 5 * exp(Z) + 2$
- $X_3^{\mathrm{cv}} = 1 + max(0, \exp(Z) - 1) = 1 + (\exp(Z) - 1)^+$

Erwartungswerte analytisch berechenbar (aus Laplace Transformierter bzw. aus 'Black-Scholes-Formel' für Call-Preise (bzw mit quadr.Ergänzung)).

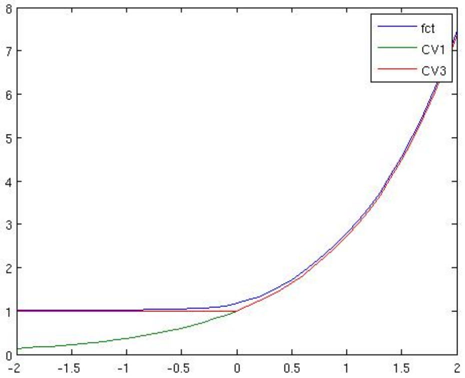
Beispiel Control Variates



Matlab: Beispiel Control Variates 1 u 2

```
CV1: log10(n)     MCmean        MCKI          KIwidth     WidthRatioCVvsPlain
          2        1.994     [1.930,2.057]    0.1269         7.181e-02
          3        1.963     [1.947,1.978]    0.0311         1.029e-01
          4        1.955     [1.951,1.960]    0.0093         1.136e-01
          5        1.956     [1.955,1.958]    0.0029         1.152e-01
Press key

Check: Identical results for affine transform control variate CV2=5*CV1-2 :

CV2: log10(n)     MCmean        MCKI          KIwidth     WidthRatioCVvsPlain
          2        1.994     [1.930,2.057]    0.1269         7.181e-02
          3        1.963     [1.947,1.978]    0.0311         1.029e-01
          4        1.955     [1.951,1.960]    0.0093         1.136e-01
          5        1.956     [1.955,1.958]    0.0029         1.152e-01
```

Beispiel: weitere Control Variate



Übersicht

1 Konfidenzintervalle bei Monte Carlo

2 Antithetic Variates

3 Control Variates

4 Importance Sampling

Importance Sampling

- Idee: Sampling dort, wo es darauf ankommt.
- Mathematisches Mittel: Maßwechsel $P_f \leftrightarrow P_g$ mit Dichte=Likelihood-Ratio.
- Sei Y Zufallsvariable mit $X := h(Y) \in L^1(P_f)$ für $h \geq 0$ und Wahrscheinlichkeitsmaß P_f, unter welchem die Verteilung von Y die Dichte f hat.
- Ziel: Berechnung von

$$E_f[X] = E_f[h(Y)] = \int_\Omega h(Y)\, dP_f = \int_{\mathbb{R}^d} h(y) f(y) dy$$

Importance Sampling

- Sei Y unter anderem Wahrscheinlichkeitsmaß P_g verteilt mit Dichte g wobei $\{g = 0\} \subseteq \{f = 0\}$.

$$\begin{aligned} E_f[h(Y)] &= \int_\Omega h(Y)\, dP_f = \int_{\mathbb{R}^d} h(y)f(y)dy \\ &= \int_{\mathbb{R}^d} h(y)\frac{f(y)}{g(y)}g(y)dy \\ &= E_g[h(Y)q(Y)]\,, \end{aligned}$$

dabei ist $\frac{dP_f}{dP_g} = \frac{f(Y)}{g(Y)} =: q(Y)$ der Likelihood-Quotient oder Radon-Nikodym Dichte von P_f bzgl. P_g.

- Ziel: Wähle g so, dass $Var_g(h(Y)q(Y)) < Var_f(h(Y))$ ist und Y unter P_g gut simuliert werden kann!
- Theoretisches Optimum: $\exists\ P_g$ mit $Var_g(h(Y)q(Y)) = 0$!

Exponential Tilting

- In wichtigen Klassen von Verteilungen kann man durch ‘Exponential Tilting‘ effizient unter einem neuen Maß aus der bekannten Familie simulieren.
- Für Normalverteilungen: Sei $Y \sim \mathcal{N}(0, Q)$ unter P_f und $Y \sim \mathcal{N}(m, Q)$ unter P_g mit $m \in \mathbb{R}^d$.
- Dann
$$q(y) = \exp\left(-x^t Q^{-1} m + \frac{1}{2} m^t Q^{-1} m\right)$$
- Diskutieren: Wie m wählen ?

Beispiel: Matlab Resultate

- $Y = Z \sim \mathcal{N}(0,1)$ unter P_f
- $h(y) = \max(0, e^y - e^3) = (e^y - e^3)^+$

Beispiel: Matlab Resultate

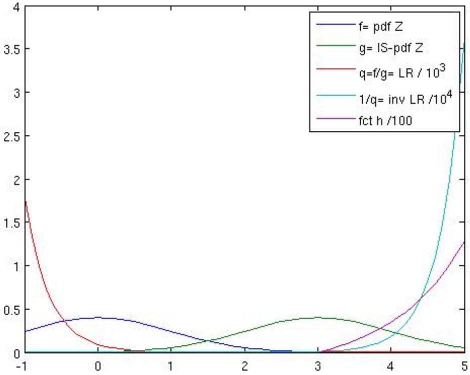
- $Y = Z \sim \mathcal{N}(0,1)$ unter P_f
- $h(y) = \max(0, e^y - e^3) = (e^y - e^3)^+$
- P_g so dass $Y \sim \mathcal{N}(3,1)$
- $q(y) = \frac{\exp(-\frac{1}{2}x^2)}{\exp(-\frac{1}{2}(x-3)^2)} = \exp(-3x + 3^2/2)$

Beispiel: Matlab Resultate

```
Variance reduction experiment  for E[max(0,exp(Z)-exp(3))]
for Z standardnormal without and with Importance Sampling

Note: True expectation known =  1.0395e-02

PlainMC:log10(n)   MCmean          MCKI                  KIwidth
          2      0.000e+00   [0.000e+00,0.000e+00]    0.000e+00
          3      4.685e-03  [-3.081e-03,1.245e-02]    1.553e-02
          4      5.460e-03   [1.128e-03,9.792e-03]    8.663e-03
          5      1.006e-02   [7.166e-03,1.295e-02]    5.779e-03


Now Importance sampling under distribution such that Z is N(3,1)-distributed:

IS-MC:  log10(n)     MCmean         MCKI                KIwidth     KIWidthRatioISvsPlain
         2         1.089e-02  [8.373e-03,1.341e-02]  5.040e-03          Inf
         3         1.059e-02  [9.816e-03,1.137e-02]  1.558e-03       1.003e-01
         4         1.044e-02  [1.019e-02,1.068e-02]  4.920e-04       5.679e-02
         5         1.034e-02  [1.026e-02,1.042e-02]  1.555e-04       2.692e-02
```

Beispiel: Importance Sampling

PDFs and Likelihood-Ratio and (scaled) Function h



Literatur

- Sheldon Ross, Simulation, Academic Press, 2006
- Paul Glasserman, Monte Carlo Methods in Financial Engineering, Springer, 2004
- Higham+Higham, Matlab Guide