MONTE CARLO METHODS
WINTER TERM 2015/16

CAROLINE LASSER

Contents

Date: January 22, 2016.

1. Random variable generation

1.1. **Direct simulation (13.10.)**

1.1.1. *Integrals.* Let $f : [0,1]^d \to \mathbb{R}$ be integrable. We want to approximate

$$I(f) = \int_{[0,1]^d} f(x)dx.$$

Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. random variables uniformly distributed on $[0,1]^d$. We will explore the approximation $\frac{1}{n}\sum_{j=1}^n f(X_j) \approx I(f)$.

1.1.2. *Mean square accuracy.* Let $Y \in L^2(\Omega)$ with $\mathbb{E}(Y) = a$, and $(Y_n)_{n\in\mathbb{N}}$ i.i.d. random variables distributed according to Y. Then, $D_n = \frac{1}{n}\sum_{j=1}^n Y_j$ satisfies

$$\mathbb{E}(D_n) = a, \qquad \sqrt{\mathbb{E}((D_n - a)^2)} = \frac{\sigma(Y)}{\sqrt{n}}.$$

Proof. $\mathbb{E}((D_n-a)^2) = \mathbb{V}(D_n) = \frac{1}{n^2}\sum_{j=1}^n \mathbb{V}(Y_j) = \frac{\mathbb{V}(Y)}{n}$, since $(Y_n)_n$ is i.i.d. □

1.1.3. *Approximating π.* Let X be uniformly distributed on $[0,1]^2$. Denote the disc by $B = \{x \in \mathbb{R}^2 \mid |x| \le 1\}$ and set $Y = \chi_B(X)$. Then, $\mathbb{E}(Y) = \pi/4$.

```
N = 1e+3; for n=1:N; x=rand(n,1); y=rand(n,1); ...
d(n)=sum(heaviside(1-x.^2-y.^2))/n; end,
loglog(1:N,abs(d-pi/4),1:N,1./sqrt([1:N]))
```

1.1.4. *Limit results.* $\lim_{n\to\infty} D_n = a$ can be meant in various ways:

1. Mean square convergence: $\lim_{n\to\infty} \mathbb{E}((D_n - a)^2) = 0$.
2. The weak law of large numbers: $\lim_{n\to\infty} P(|D_n - a| \ge \varepsilon) = 0$ for all $\varepsilon > 0$.
3. The strong law of large numbers: $\lim_{n\to\infty} D_n = a$ almost everywhere.
4. The central limit theorem: The standardised sum $D_n^* = \frac{\sqrt{n}}{\sigma(Y)}(D_n - a)$ satisfies

$$\lim_{n\to\infty} P(D_n^* \in]-\infty, u]) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^u e^{-x^2/2}dx$$

uniformly in $u \in \mathbb{R}$.

1.1.5. *Empirical variance.* Let $Y \in L^2(\Omega)$ und $(Y_n)_{n\in\mathbb{N}}$ i.i.d. random variables distributed according to Y. The empirical variance

$$V_n = \frac{1}{n-1}\sum_{j=1}^n (Y_j - D_n)^2$$

satisfies $\mathbb{E}(V_n) = \mathbb{V}(Y)$ and $\lim_{n\to\infty} V_n = \mathbb{V}(Y)$ almost everywhere. That is, V_n is an unbiased and strongly consistent estimator for $\mathbb{V}(Y)$.

Proof. From

$$V_n = \frac{1}{n-1}\sum_{j=1}^n (Y_j - a)^2 - \frac{n}{n-1}(D_n - a)^2$$

we deduce $\mathbb{E}(V_n) = (\frac{n}{n-1} - \frac{1}{n-1})\mathbb{V}(Y) = \mathbb{V}(Y)$. Applying the strong law twice, we get $\lim_{n\to\infty} \frac{n}{n-1}(D_n - a)^2 = 0$ a.e. and $\lim_{n\to\infty} V_n = \mathbb{V}(Y)$ a.e. □

1.1.6. *Recommended reading.* [MNR, 3.1–3.5]

1.2. **Inversion and rejection method (20.10.)**

1.2.1. *The inversion principle.* Let $F : \mathbb{R} \to [0, 1]$ be a continuous cumulative distribution function. Then,

$$F^{-1} :]0,1[\to \mathbb{R}, \quad u \mapsto \inf\{x \in \mathbb{R} \mid F(x) = u\}$$

is well-defined. Moreover, if U is uniformly distributed on $[0, 1]$, then $F^{-1}(U)$ has cumulative distribution function F.

Proof. Since $F^{-1}(u) \le x$ iff $u \le F(x)$, $P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$. $\square$

1.2.2. *Exponential generator by inversion.* The density of the exponential distribution is $f(x) = \mu e^{-\mu x}\chi_{[0,\infty[}(x)$. Then, $F(x) = (1 - e^{-\mu x})\chi_{[0,\infty[}(x)$ and $F^{-1}(u) = -\frac{1}{\mu}\ln(1 - u)$.

```
mu = 2; n = 1e+3; u = rand(n,1); y = -log(u)/mu; a = sum(y)/n
```

1.2.3. *Property of densities.* Let f be a probability density on $\mathbb{R}^d$ and $X \sim f$. Let U be uniformly distributed on $[0, 1]$, and let $c > 0$. If X and U are independent, then $(X, cUf(X))$ is uniformly distributed on

$$A = \left\{(x, u) \mid x \in \mathbb{R}^d, 0 \le u \le cf(x)\right\}.$$

Proof. Let $B \in \mathcal{B}(\mathbb{R}^{d+1})$, $B \subseteq A$. Since $\mathrm{vol}(A) = c$, it enough to observe

$$P((X, cUf(X)) \in B) = \int_{\mathbb{R}^d}\int_{\{u|(x,u)\in B\}} \frac{du}{cf(x)} f(x)dx = \frac{1}{c}\int_B dudx. \qquad \square$$

1.2.4. *The rejection principle.* Let (X_n) be i.i.d. with values in $\mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$ such that $P(X_1 \in A) > 0$. Let Y be the first X_n taking values in A. Then, for all $B \in \mathcal{B}(\mathbb{R}^d)$,

$$P(Y \in B) = P(X_1 \in A \cap B)/P(X_1 \in A).$$

Proof. We set $p = P(X_1 \in A)$ and use the geometric series at the end of

$$P(Y \in B) = \sum_{n=1}^{\infty} P(X_1 \notin A, \ldots, X_n \in A \cap B) = \sum_{n=1}^{\infty}(1 - p)^{n-1}P(X_1 \in A \cap B).$$

1.2.5. *The rejection method.* Let f, g be densities on $\mathbb{R}^d$ and $c > 0$ such that

$$f(x) \le cg(x) \qquad (x \in \mathbb{R}^d).$$

Generate $X \sim g$ and independently U uniformly distributed on $[0, 1]$. If $cUg(X) > f(X)$, then repeat the same, otherwise return X. Then, $X \sim f$.

Proof. Before exit, the $(X, cUg(X))$ are uniformly distributed below cg. After exit, $cUg(X) \le f(X)$, and the rejection principle implies that $(X, cUg(X))$ is uniformly distributed below f. Then, for all $B \in \mathcal{B}(\mathbb{R}^d)$,

$$P(X \in B) = P((X, cUg(X)) \in \{(x, u) \mid x \in B, 0 \le u \le f(x)\}) = \int_B f(x)dx. \quad \square$$

1.2.6. *The number of iterations.* The acceptance probability is

$$p = P(f(X) \geq cUg(X)) = \int_{\mathbb{R}^d} \frac{f(x)}{cg(x)} g(x) dx = \frac{1}{c}.$$

The number of iterations N satisfies $p(N = k) = (1-p)^{k-1}p$ for $k \geq 1$. N is geometric, and we have $\mathbb{E}(N) = 1/p = c$, $\mathbb{V}(N) = (1-p)/p^2 = c^2 - c$.

1.2.7. *Recommended reading.* [De, II.2 & II.3]

1.3. **Uniform pseudorandom numbers (27.10.)**

1.3.1. *Modular arithmetic.* Let $a, b \in \mathbb{Z}$ and $m \in \mathbb{N}_{>0}$. Then, there exist unique $q, r \in \mathbb{Z}$ such that $a = qm + r$ with $0 \leq r < m$. We say $a = b \bmod m$, if m divides $a - b$ without remainder. Congruence modulo m is an equivalence relation on $\mathbb{Z}$.

1.3.2. *Linear congruential generators.* Let $a, c, x \in \mathbb{Z}$, $m \in \mathbb{N}_{>0}$ and define the associated linear congruential random number generator (LCRNG) with seed x as

$$f(x) = (ax + c) \bmod m.$$

```
x(1) = 0; % x = 1e-16; ...
for n=1:50, x(n+1) = mod(5*x(n)+1,8); end, ...
hold on, plot([1:51],x,'-o')
```

1.3.3. *Maximal period.* There exists an equivalent condition on a, c, m such that the linear congruential random number generator has its maximal period m for all seeds x.

1.3.4. *Uniform distribution on* $[0, 1[$. A linear congruential random number generator with maximal period m has as its range $0, 1, \ldots, m-1$. Division by m produces a sequence in $[0, 1[$.

1.3.5. *Fibonacci generators.* Let $a, x \in \mathbb{Z}^n$ such that $|\text{supp}(a)| = 2$, $c \in \mathbb{Z}$, $m \in \mathbb{N}_{>0}$ and define the associated Fibonacci generator with seed x as

$$f(x) = (a \cdot x + c) \bmod m$$

The choice $n = 2$, $a = (1, 1)$, $c = 0$ and $x = (0, 1)$ generates the Fibonacci sequence.

```
% Caricature of a Marsaglia Fibonacci generator
m = 97; x = randi(m,55,1); plot([1:55],x,'-o'), hold on, ...
for n=56:110, x(n) = mod(x(n-23)+ x(n-54),m); end,...
plot([1:55],x(56:end),'r-*')
```

1.3.6. *Mersenne twister.* The Mersenne twister is a linear generator over the binary field $\mathbb{F}_2$. Let $A \in \mathbb{F}_2^{k \times k}$ the transition matrix, and $B \in \mathbb{F}_2^{w \times k}$ the output transformation matrix. Then the ith output $u_i \in [0, 1[$ is constructed by

$$x_i = Ax_{i-1}, \qquad y_i = Bx_i, \qquad u_i = \sum_{l=1}^{w} y_{i,l-1} 2^{-l}.$$

The period length is the Mersenne prime $2^k - 1$ with $k = 19937$.

1.3.7. *Binomial chi-square test.* We test whether n independent coin tosses come from a binomial distribution with parameter p. Let Y be the number of heads and

$$V = \left(\frac{Y - np}{\sqrt{np(1-p)}} \right)^2 = \frac{(Y - np)^2}{np} + \frac{(n - Y - n(1-p))^2}{n(1-p)}.$$

By the central limit theorem, V is approximately the square of a standard normally distributed random variable and thus χ_1^2-distributed.

```
n = 100; for k=1:n, x = randi(2,n,1); y = sum(x-1); ...
v(k) = (y-0.5*n)^2/(n*0.25); end, subplot(2,1,1), hist(v),  ...
x = linspace(0,max(v)); y = chi2pdf(x,1); ...
subplot(2,1,2), plot(x,y,'r-*')
```

1.3.8. *Multinomial chi-square test.* We test whether n independent experiments come from a multinomial distribution with parameters $p_1, \ldots, p_k$. Let Y_j be the number of outcomes with property j. Then,

$$V = \sum_{j=1}^{k} \frac{(Y_j - np_j)^2}{np_j}$$

is approximately χ^2_{k-1} distributed.

1.3.9. *Glivenko–Cantelli Theorem.* Let $X_1, \ldots, X_n$ be i.i.d. with cumulative distriution function $F : \mathbb{R} \to [0, 1]$. Then, the empirical distribution function

$$F_n(x) = \tfrac{1}{n}|\{j = 1, \ldots, n : X_j \le x\}|, \qquad x \in \mathbb{R},$$

satisfies $\lim_{n\to\infty} \|F_n - F\|_\infty = 0$ almost surely.

```
% Kolmogorov-Smirnov test
n = 100; y = rand(n,1); cdfplot(y), hold on, [f,x] = ecdf(y); ...
plot(x,x,'-r'), d = max(abs(f-x))
% Asymptotics
N = [1:2e+3]; for n=N, y = rand(n,1); [f,x] = ecdf(y); ...
d(n) = max(abs(f-x)); end, semilogy(N,d,'b',N,1./sqrt(N),'r')
```

1.3.10. *Recommended reading.* [SM, Chapter 1.3 & Appendix A]

2. Variance reduction

2.1. **Antithetic and control variates (03.11.)** In this lecture, all random variables are square integrable.

2.1.1. *Symmetric random vectors.* A random vector X is symmetrically distributed with respect to $\mu \in \mathbb{R}^d$, if

$$P(X \in \mu + A) = P(X \in \mu - A), \qquad A \in \mathcal{B}(\mathbb{R}^d),$$

that is, $X \sim 2\mu - X$. Examples are the uniform distribution on $[0,1]^d$ with $\mu = (0.5, \dots, 0.5)$ and the standard normal distribution with $\mu = 0$.

2.1.2. *Even and odd part.* Let X be symmetric with respect to μ and consider for $f : \mathbb{R}^d \to \mathbb{R}$ the even and the odd part, $f_\pm(x) = \frac{1}{2}\left(f(x) \pm f(2\mu - x)\right)$, $x \in \mathbb{R}^d$. Then, $\mathbb{E}(f_+(X)) = \mathbb{E}(f(X))$, $\mathbb{E}(f_-(X)) = 0$, and

$$\mathbb{V}(f(X)) = \mathbb{V}(f_+(X)) + \mathbb{V}(f_-(X)).$$

Proof. $\mathbb{E}(f_+(X)f_-(X)) = \mathbb{E}(f_+(2\mu - X)f_-(2\mu - X)) = -\mathbb{E}(f_+(X)f_-(X))$. □

2.1.3. *Antithetic sampling.* Let X be symmetric with respect to μ. We directly simulate $a = \mathbb{E}(f(X))$ by $D_n = \frac{1}{n}\sum_{j=1}^n f(X_j)$ and $\widetilde{D}_n = \frac{1}{n}\sum_{j=1}^n f_+(X_j)$. Then,

$$\sqrt{\frac{\mathbb{E}((D_{2n} - a)^2)}{\mathbb{E}((\widetilde{D}_n - a)^2)}} = \frac{\sigma(f(X))}{\sqrt{2}\sigma(f_+(X))} \geq \frac{1}{\sqrt{2}} \approx 0.7.$$

2.1.4. *Example.* We consider $f(x) = \exp(\sqrt{x})$ and $a = \int_0^1 f(x)dx = 2$.

```
N = [1:1e+3]; for n = N, x = rand(n,1); y = exp(sqrt(x)); ...
d(n) = mean(y); ya = (exp(sqrt(x)) + exp(sqrt(1-x)))/2; ...
da(n) = mean(ya); end, semilogy(N,abs(d-2),N,abs(da-2),N,1./sqrt(N))
```

2.1.5. *Control variates.* Let Y, Z be random variables and $b \in \mathbb{R}$. Then, $Y_b = Y - b(Z - \mathbb{E}(Z))$ satisfies $\mathbb{E}(Y) = \mathbb{E}(Y_b)$ and

$$\min_{b \in \mathbb{R}} \mathbb{V}(Y_b) = \mathbb{V}(Y_{b^*}), \qquad b^* = \mathrm{Cov(Y,Z)}/\mathbb{V}(Z).$$

Proof. $\mathbb{V}(Y_b) = \mathbb{V}(Y_b - \mathbb{E}(Y)) = \mathbb{V}(Y) - 2b\mathrm{Cov(Y,Z)} + b^2\mathbb{V}(Z)$. □

2.1.6. *Example.* We reconsider $\int_0^1 \exp(\sqrt{x})dx = 2$ and use $Z = \widetilde{f}(X)$ with $\widetilde{f}(x) = 1 + (e-1)x$ and X uniformly distributed on $[0,1]$ as control variate. The corresponding direct simulation is

$$\widetilde{D}_n = b\mathbb{E}(\widetilde{f}(X)) + \frac{1}{n}\sum_{j=1}^n (f - b\widetilde{f})(X_j).$$

```
b = 1; % b= 0.89; ...
for n=N, x = rand(n,1); yc = exp(sqrt(x)) - b - b*(exp(1)-1)*x; ...
dc(n) = mean(yc) + b*(exp(1) + 1)/2; end, ...
hold on, semilogy(N,abs(dc-2),'m')
```

2.1.7. *Empirical covariance.* For estimating the optimal $b^* = \mathrm{Cov}(Y,Z)/\mathbb{V}(Z)$ one might use the empirical covariance, that is,

$$b^* \approx \frac{\sum_{j=1}^{n}(Y_j - D_n(Y))(Z_j - D_n(Z))}{\sum_{j=1}^{n}(Z_j - D_n(Z))^2},$$

where $(Y_1,\ldots,Y_n)$ and $(Z_1,\ldots,Z_n)$ are i.i.d. according to Y and Z, respectively.

2.1.8. *Recommended reading.* [MNR, 5.1 & 5.2]

2.2. **Stratified and importance sampling (17.11.)**

2.2.1. *Stratification.* Let X be a random vector and $A_1, \dots, A_m \subseteq \mathbb{R}^d$ a disjoint partition of its range with $p_j = P(X \in A_j) > 0$ for all j. Denote $a = \mathbb{E}(f(X))$, $\sigma^2 = \mathbb{V}(f(X))$, and $Q_j = P(X \in \cdot \mid X \in A_j)$, $a_j = \int_{\mathbb{R}^d} f(x) dQ_j(x)$, $\sigma_j^2 = \int_{\mathbb{R}^d} (f(x) - a_j)^2 dQ_j(x)$. Then,

$$a = \sum_{j=1}^m p_j a_j, \quad \sigma^2 = \sum_{j=1}^m p_j \left(\sigma_j^2 + (a_j - a)^2\right).$$

Proof. We have $a = \sum_{j=1}^m \mathbb{E}(f(X)\chi_{A_j}(X)) = \sum_{j=1}^m p_j a_j$ and

$$\sigma^2 = \mathbb{E}(f(X)^2) - \mathbb{E}(f(X))^2 = \sum_{j=1}^m p_j(\sigma_j^2 + a_j^2 - a^2) = \sum_{j=1}^m p_j \left(\sigma_j^2 + (a_j - a)^2\right),$$

since $\sum_{j=1}^m p_j(a_j - a)^2 = \sum_{j=1}^m p_j(a_j^2 - 2a_j a + a^2) = \sum_{j=1}^m p_j(a_j^2 - a^2)$. ☐

2.2.2. *Stratified sampling.* Let $X_i = (X_i^{(1)}, \dots, X_i^{(m)})$, $i \in \mathbb{N}$, be an i.i.d. sequence such that $X_i^{(j)} \sim Q_j$. Let $n \in \mathbb{N}$ and $n_j = p_j n$. Then,

$$S_n = \sum_{j=1}^m p_j \cdot \frac{1}{n_j} \sum_{i=1}^{n_j} f(X_i^{(j)})$$

satisfies $\mathbb{E}(S_n) = a$ and $\mathbb{V}(S_n) = \sum_{j=1}^m p_j^2 \sigma_j^2 / n_j = \frac{1}{n} \sum_{j=1}^m p_j \sigma_j^2$

2.2.3. *Comparison with direct simulation.* Let $X_1, \dots, X_n$ be i.i.d. according to X, and $D_n = \frac{1}{n} \sum_{j=1}^n f(X_j)$. Then,

$$\sqrt{\mathbb{E}((D_n - a)^2)} = \frac{\sigma}{\sqrt{n}} \geq \frac{1}{\sqrt{n}} = \sqrt{\mathbb{E}((S_n - a)^2)}$$

with equality iff $a_1 = \dots = a_m = a$.

2.2.4. *Univariate integration.* Let $f : [0, 1] \to \mathbb{R}$ and $z_j = \frac{j}{m}$. We set $A_j = [z_{j-1}, z_j]$ and $n = k \cdot m$ such that $p_j = \frac{1}{m}$ and $n_j = k$. Consider $X_i^{(j)} = z_{j-1} + (z_j - z_{j-1}) U_i^{(j)}$ for i.i.d. random variables $U_i^{(j)}$ uniformly distributed on $[0, 1]$. Then,

$$S_n = \frac{1}{k \cdot m} \sum_{j=1}^m \sum_{i=1}^k f(X_i^{(j)})$$

```
K = [1:1e+3]; m = 100; ...
for k=K, u = (rand(k,m) + repmat([0:1:m-1],k,1))/m; ...
y = exp(sqrt(u)); s(k) = mean(y(:)); ...
x = rand(k*m,1); y = exp(sqrt(x)); d(k) = mean(y); end,
semilogy(K,abs(s-2),'g',K,abs(d-2),'b', K,1./sqrt(K),'r')
```

2.2.5. *Importance sampling.* Let $\Omega \subseteq \mathbb{R}^d$ and $\mu, \omega : \Omega \to [0,\infty)$ be two probability densities with μ/ω well defined. We simulate $a = \int_\Omega f(x)\mu(x)dx$ by

$$I_n = \frac{1}{n}\sum_{j=1}^{n} f(X_j)\frac{\mu(X_j)}{\omega(X_j)},$$

where $X_1,\ldots,X_n$ are i.i.d. according to ω. We have $\mathbb{E}(I_n) = a$ and

$$\sqrt{\mathbb{E}((I_n - a)^2)} = \sqrt{\frac{1}{n}\int_\Omega \left(\frac{f(x)\mu(x)}{\omega(x)} - a\right)^2 \omega(x)dx}.$$

We therefore seek ω such that $f\mu/\omega$ is almost constant.

2.2.6. *Univariate integration.* We consider $\Omega = [0,1]$, $\mu = 1$ and $f(x) = x^{\alpha-1}e^{-x}$ for $\alpha > 0.5$. We compare $\omega = 1$ and $\omega(x) = \alpha x^{\alpha-1}$. We use the inversion method: We have $F(u) = u^\alpha$ and $F^{-1}(x) = x^{1/\alpha}$.

```
a = 0.51; N = [1:1e+3]; for n=N,
x = rand(n,1); y = x.^(a-1).*exp(-x); d(n) = mean(y); ...
u = rand(n,1); y = exp(-u.^(1/a))/a; i(n) = mean(y); end,...
plot(N,d,N,i), r = i(end); pause, ...
semilogy(N,abs(d-r),N,abs(i-r),N,1./sqrt(N))
```

2.2.7. *Recommended reading.* [MNR, 5.3 & 5.4], [NW10, Chapter 17.2]

3. Markov Chains

3.1. **Countable Markov Chains (24.11.)**

3.1.1. *Countable Markov chains.* A sequence of random variables $(X_n)_{n\geq 0}$ with values in a countable set E is a *Markov chain*, if for all $i_0, \ldots, i_{n-1}, i, j \in E$

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j \mid X_n = i),$$

whenever both conditional probabilities are well defined. We consider only *homogeneous* Markov chains, where the above is independent of n for all (i, j).

3.1.2. *Example: Ehrenfest diffusion.* Consider M molecules in two boxes and denote by X_n the number of molecules in the left box. Let $X_0 = M$ and

$$P(X_{n+1} = i + 1 \mid X_n = i) = (M - i)/M, \quad i < M,$$
$$P(X_{n+1} = i - 1 \mid X_n = i) = i/M, \quad i > 0,$$

and $P(X_{n+1} = j \mid X_n = i) = 0$ if $j \neq i \pm 1$. Then, $(X_n)_{n\geq 0}$ is a Markov chain with values in $\{0, \ldots, M\}$.

```
M = 100; x(1) = M; N = [1:1e+5]; % N = [1:100] ...
for n=N, u = randi(M); if u <= x(n), x(n+1) = x(n)-1; ...
else, x(n+1) = x(n)+1; end, end, plot(N,x(1:end-1),'-')
```

3.1.3. *Transition matrix.* The transition probabilities are defined by

$$p_{ij} := P(X_{n+1} = j \mid X_n = i), \qquad i, j \in E.$$

The transition matrix $P = (p_{ij})$ is stochastic, since $p_{ij} \geq 0$ and $\sum_{j\in E} p_{ij} = 1$. In particular, 1 is eigenvalue of P and $|\lambda| \leq 1$ for all eigenvalues λ of P.

Proof. 1 is eigenvalue, since $v = (1, \ldots, 1)^T$ satisfies $Pv = v$. For λ an eigenvalue with eigenvector v with $\|v\|_\infty = 1$, we have

$$|\lambda| = \|\lambda v\|_\infty = \|Pv\|_\infty \leq \max_i \sum_j |p_{ij} v_j| \leq \max_i \sum_j p_{ij} = 1. \qquad \square$$

3.1.4. *Example: Ehrenfest diffusion.* The transition matrix of a $M = 2$ Ehrenfest chain satisfies

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \end{pmatrix}, \qquad P^2 = \begin{pmatrix} 0.5 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \end{pmatrix}, \qquad P^3 = P,$$

so that $P^{2k} = P^2$ and $P^{2k+1} = P$ for all $k \geq 1$. Hence, for odd n there is no return to the initial state.

3.1.5. *Stationary distribution.* A probability distribution on E is a vector μ with $\mu \geq 0$ (that is, $\mu_i \geq 0$ for all $i \in E$) and $\|\mu\|_1 = 1$. It is called *stationary* with respect to a stochastic matrix P if it is in global balance, that is,

$$\mu^T P = \mu^T.$$

A Markov chain with $X_0 \sim \mu$, where μ is stationary with respect to P, satisfies $X_n \sim \mu$ for all n.

3.1.6. *Ehrenfest chain.* The Ehrenfest chain has the stationary distribution

$$\mu_i = \binom{M}{i} 2^{-M}, \qquad i = 0, \dots, M,$$

since for all $j = 1, \dots, M-1$

$$\begin{aligned}(\mu^T P)_j &= \mu_{j-1} p_{j-1,j} + \mu_{j+1} p_{j+1,j} \\ &= 2^{-M} \left(\binom{M}{j-1} \frac{M-(j-1)}{M} + \binom{M}{j+1} \frac{j+1}{M} \right) = \mu_j.\end{aligned}$$

```
M = 100; x(1) = M; N = [1:1e+5]; ...
for n=N, u = randi(M); if u <= x(n), x(n+1) = x(n)-1; ...
else, x(n+1) = x(n)+1; end, end, ...
b = histc(x,[0:M]); bar([0:M],b/N(end),'histc'); ...
mu = factorial(M)./factorial([0:M])./factorial(M-[0:M])/2^M; ...
hold on, plot([0:M],mu,'-r')
```

3.1.7. *Contraction argument.* Let μ, ν be probability distributions, μ stationary with respect to P. Then, for all $n \in \mathbb{N}$,

$$\left\| \nu^T P^n - \mu^T \right\|_1 = \left\| (\nu - \mu)^T P^n \right\|_1 \le \sup \left\{ \left\| w^T P^n \right\|_1 : \sum_{j \in E} w_j = 0 \right\}.$$

Dobrushin's ergodicity coefficient of a stochastic matrix P is defined as

$$c(P) := \sup \left\{ \left\| w^T P \right\|_1 / \|w\|_1 : w \neq 0, \sum_{j \in E} w_j = 0 \right\} \in [0, 1].$$

3.1.8. *Finite positive matrices.* A finite stochastic matrix P with $p_{ij} > 0$ for all i, j satisfies $c(P) < 1$.

Proof. Recall that two real numbers $a, b \neq 0$ of opposite sign satisfy $|a+b| < |a|+|b|$. For $w \neq 0$ with $\sum_j w_j = 0$ we have

$$\|w^T P\|_1 = \sum_j |\sum_i w_i p_{ij}| < \sum_j \sum_i |w_i| p_{ij} = \|w\|_1.$$

For finite P, the supremum in $c(P)$ is a maximum. □

3.1.9. *Recommended reading.* [S, §3–4], [W, §4]

3.2. **More on Markov Chains (01.12.)**

3.2.1. *Ergodic theorem.* For a stochastic matrix P the following are equivalent:

(1) There exists $n \in \mathbb{N}$ such that $c(P^n) < 1$.
(2) There exists a probability density μ such that $\lim_{n\to\infty} \nu^T P^n = \mu^T$ uniformly for all $\nu \in \ell^1(E)$.

In particular, such a μ is stationary with respect to P and unique.

Proof. We only discuss stationarity and uniqueness: $\mu^T P = \lim_{n\to\infty} \nu^T P^{n+1} = \mu^T$. Let $\mu, \widetilde{\mu}$ satisfy (2). Then, $\widetilde{\mu}^T = \lim_{n\to\infty} \mu^T P^n = \mu^T$. □

3.2.2. *Law of large numbers.* Let $(X_n)_{n\geq 0}$ a Markov process with transition matrix P satisfying the ergodic theorem. Then, for all initial distributions ν,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(X_i) = \mathbb{E}_\mu(f) \quad \text{in} \quad L^2(\mathbb{P}_\nu)$$

for all μ-summable functions $f : E \to \mathbb{R}$ with $\sum_{x\in E} f_x^2 < \infty$.

3.2.3. *Quadratic deviation for point evaluation.* For $f = \chi_{\{x\}}$, $x \in E$, we have $\mathbb{E}_\mu(f) = \mu_x$ and

$$\mathbb{E}\Big(\big(\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}_\mu(f)\big)^2\Big) = \frac{1}{n^2}\sum_{i,j=1}^{n} \mathbb{E}\big(\,(f(X_i) - \mu_x))\,(f(X_j) - \mu_x)\,\big)$$

$$= \frac{1}{n^2}\sum_{i,j=1}^{n} \big(\nu_{i,j}(x) - \mu_x(\nu^T P^i)_x - \mu_x(\nu^T P^j)_x + \mu_x^2\big)$$

with $\nu_{i,j}(x) = \mathbb{P}_\nu(X_i = x, X_j = x)$ the two-sided marginal distributions.

3.2.4. *Cesàro means.* A convergent sequence $(c_n)_{n\geq 1}$ satisfies

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} c_i = \lim_{n\to\infty} c_n.$$

Consequently, the ergodic theorem implies $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(\nu^T P^i)_x = \mu_x$. One can also derive

$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{i,j=1}^{n} \nu_{i,j}(x) = \mu_x^2,$$

using that $\nu_{i,j}(x) = (\nu^T P^i)_x (e_x^T P^{j-i})_x$ for $j \geq i$.

3.2.5. *Detailed balance.* A probability distribution μ and a stochastic matrix P satisfy the detailed balance condition, if

$$\mu_i p_{ij} = \mu_j p_{ji}, \qquad i, j \in E.$$

Detailed balance implies global balance. If $X_0 \sim \mu$, then detailed balance means $P(X_1 = j, X_0 = i) = P(X_1 = i, X_0 = j)$ for all i, j with $\mu_i, \mu_j > 0$.

Proof. We have $(\mu^T P)_j = \sum_i \mu_i p_{ij} = \sum_i \mu_j p_{ji} = \mu_j$ and $\mu_i P(X_1 = j \mid X_0 = i) = \mu_j P(X_1 = i \mid X_0 = j)$. □

3.2.6. *Ehrenfest chain.* The Ehrenfest chain satisfies detailed balance, since

$$\mu_{j-1}p_{j-1,j} = 2^{-M}\binom{M}{j-1}\frac{M-(j-1)}{M} = 2^{-M}\binom{M}{j}\frac{j}{M} = \mu_j p_{j,j-1}.$$

3.2.7. *Reversibility.* Let $\mu > 0$ be stationary with respect to P and $X_0 \sim \mu$. Consider the stochastic matrix $Q = (q_{ij})$ with

$$q_{ij} := P(X_n = j \mid X_{n+1} = i) = \frac{P(X_{n+1} = i \mid X_n = j)P(X_n = j)}{P(X_{n+1} = i)} = \frac{p_{ji}\mu_j}{\mu_i}.$$

If detailed balance holds, then $P = Q$, and the Markov chain is called reversible with respect to μ.

3.2.8. *Self-adjointness.* If μ and P are in detailed balance, then

$$\langle x, Py\rangle_\mu = \sum_i x_i(Py)_i\mu_i = \sum_{i,j} x_i p_{ij} y_j \mu_i = \sum_{i,j} x_i p_{ji} y_j \mu_j = \langle Px, y\rangle_\mu$$

for all $x, y \in \ell^\infty(E)$. Conversely, if $\langle x, Py\rangle_\mu = \langle Px, y\rangle_\mu$ for all $x, y \in \ell^\infty(E)$, then $\langle e_k, Pe_l\rangle_\mu = \langle Pe_k, e_l\rangle_\mu$ for all $k, l \in E$, that is, μ and P are in detailed balance.

3.2.9. *Stochastic operators.* A linear operator $P : L^1(X,\nu) \to L^1(X,\nu)$ is called *stochastic*, if $\|Pf\|_1 \le \|f\|_1$ for all f and

$$Pf \ge 0, \quad \|Pf\|_1 = \|f\|_1 \quad \text{for all } f \ge 0.$$

A stochastic matrix is a stochastic operator for $L^1(E,\nu)$ with ν the counting measure on E.

3.2.10. *Transition kernels with density.* A probability density is a $\rho \in L^1(X,\nu)$ with $\rho \ge 0$ and $|\rho\|_1 = 1$. Let $p : X \times X \to [0,\infty[$ such that $p(x,\cdot)$ is a probability density for all $x \in X$. Then,

$$(Pf)(y) = \int_X f(x)p(x,y)d\nu(x), \qquad y \in X,$$

is a stochastic operator $L^1(X,\nu) \to L^1(X,\nu)$.

Proof. $\|Pf\|_1 = \int_X |\int_X f(x)p(x,y)d\nu(x)|d\nu(y) \le \|f\|_1$ for all $f \in L^1(X,\nu)$. ☐

3.2.11. *Ergodicity coefficient.* For a stochastic operator P, we define the ergodicity coefficient as

$$c(P) = \sup\left\{\|Pf\|_1/\|f_1\| : f \ne 0, \int_X f(x)d\nu(x) = 0\right\} \in [0,1]$$

and obtain the ergodic theorem of §3.2.1 with (2) generalized to: There exists a probability density ρ_* such that $\lim_{n\to\infty} P^n\rho = \rho_*$ uniformly for all $\rho \in L^1(X,\nu)$.

3.2.12. *Recommended reading.* [S, §3 & §5], [W, §4]

4. Markov Chain Monte Carlo

4.1. **Gibbs Sampling (08.12.)**

4.1.1. *Gibbs fields.* Let S be a finite set, and consider the configuration space $X = \prod_{s\in S} X_s$ with each X_s a finite set. Then,

$$Z = \sum_{x\in X} \exp(-\beta H(x)), \qquad \Pi(x) = Z^{-1}\exp(-\beta H(x))$$

define the *partition function* and the *Gibbs field* for the *inverse temperature* $\beta > 0$ and the *energy function* $H : X \to \mathbb{R}$. Note that Π is a strictly positive probability distribution on X.

4.1.2. *Ising model.* Let $S = \{1, \dots, m\}^2$, $X_s = \{-1, 1\}$ for $s \in S$. Set

$$H(x) = -J \sum_{\|s-t\|_1=1} x_s x_t, \qquad x \in X,$$

where in the case $J > 0$ configurations pointing in the same direction are favoured (*ferromagnets*), while $J < 0$ favours opposite directions (*antiferromagnets*).

4.1.3. *Local characteristics.* For $I \subset S$ define a *local characteristic* of Π by

$$\Pi_I(x,y) = \begin{cases} \Pi(X_I = y_I \mid X_{S\setminus I} = x_{S\setminus I}), & y_{S\setminus I} = x_{S\setminus I}, \\ 0, & \text{otherwise.} \end{cases}$$

Π_I is a stochastic matrix, since for all $x \in X$,

$$\sum_{y\in X} \Pi_I(x,y) = \sum_{y: y_{S\setminus I} = x_{S\setminus I}} \Pi(X_I = y_I \mid X_{S\setminus I} = x_{S\setminus I}) = 1.$$

For the single-site characteristics of the Ising model, we compute

$$\begin{aligned} \Pi(X_s = x_s \mid X_t = x_t, t \neq s) &= \frac{\exp(\beta J x_s \sum_{\|s-t\|_1=1} x_t)}{\exp(-\beta J \sum_{\|s-t\|_1=1} x_t) + \exp(\beta J \sum_{\|s-t\|_1=1} x_t)} \\ &= \big(1 + \exp(-2\beta J x_s \sum_{\|s-t\|_1=1} x_t)\big)^{-1}. \end{aligned}$$

4.1.4. *Reversibility.* A Gibbs field Π is reversible and thus stationary for its local characteristics.

Proof. We check detailed balance for $x, y \in X$ with $x_{S\setminus I} = y_{S\setminus I}$ and observe

$$\begin{aligned} \Pi(x)\Pi_I(x,y) &= \exp(-\beta H(x)) \frac{\exp(-\beta H(y_I, x_{S\setminus I}))}{\sum_{z_I} \exp(-\beta H(z_I, x_{S\setminus I}))} \\ &= \exp(-\beta H(x_I, y_{S\setminus I})) \frac{\exp(-\beta H(y))}{\sum_{z_I} \exp(-\beta H(z_I, y_{S\setminus I}))} = \Pi(y)\Pi_I(y,x). \end{aligned}$$

4.1.5. *Gibbs sampler.* We enumerate $S = \{1, \dots, \sigma\}$ and set

$$P = \Pi_{\{1\}} \cdots \Pi_{\{\sigma\}}.$$

Then, $\lim_{n\to\infty} \nu^T P^n = \Pi^T$ uniformly for all initial distributions ν.

Proof. Since P is a finite positive matrix, we have $c(P) < 1$ and thus convergence to the stationary distribution Π. □

4.1.6. *Numerical experiment.* We consider the Ising model for a square lattice torus.

```
m = 20; b = 0.8; J = -1; x = 2*randi(2,m,m)-3; ...
xx = [0,x(end,:),0; x(:,end),x,x(:,1);0,x(1,:),0]; ...
kk = 1, subplot(2,4,kk), pcolor(x), drawnow, ...
pp = 1./(1+exp(-2*b*J*[-4:2:4])); p([1,3,5,7,9]) = pp; ...
for k = 1:3e+3, for i = 1:m, for j = 1:m, u = rand; ...
ls = xx(i,j+1) + xx(i+2,j+1) + xx(i+1,j) + xx(i+1,j+2); ...
if (u <= p(ls+5)), x(i,j)=1; else x(i,j)=-1; end, ...
if (i==1|i==m|j==1|j==m), ...
xx = [0,x(end,:),0; x(:,end),x,x(:,1);0,x(1,:),0]; end, ...
end, end, ...
if (k==1|k==3|k==10|k==30|k==100|k==1e+3|k==3e+3), kk=kk+1, ...
subplot(2,4,kk), pcolor(x), drawnow, end, end
```

4.1.7. *Recommended reading.* [W, §3 & §5], [MNR, 6.3]

4.2. **Annealing and Metropolis Sampling (15.12.)**

4.2.1. *Algorithmic interpretation.* Consider a finite space $X = \prod_{s=1}^{\sigma} X_s$, the Gibbs sampler

$$P = \Pi_{\{1\}} \cdots \Pi_{\{\sigma\}}$$

and an initial distribution ν. Then, $\nu^T P$ means the following: Pick an initial configuration $x = (x_1, x')$ according to ν. Update x_1 according to $\Pi_{\{1\}}((x_1, x'), (\cdot, x'))$. Visit all the dimensions. For the update of the last component of $y = (y', y_n)$ use $\Pi_{\{n\}}((y', y_n), (y', \cdot))$.

4.2.2. *Positivity.* The Gibbs sampler $P = \Pi_{\{1\}} \cdots \Pi_{\{\sigma\}}$ is a positive matrix.

Proof. If $\sigma = 2$, then for all $x, y \in X$

$$P(x, y) = \sum_{z \in X} \Pi_{\{1\}}(x, z)\Pi_{\{2\}}(z, y) = \Pi_{\{1\}}(x, (y_1, x_2))\Pi_{\{2\}}((y_1, x_2), y) > 0. \qquad \square$$

4.2.3. *Limits.* Let $\Pi = \Pi^\beta$ be a Gibbs field on a finite set X and $M \subseteq X$ the set of its maxima. Then, for all $x \in X$,

$$\lim_{\beta \to 0} \Pi^\beta(x) = |X|^{-1}, \qquad \lim_{\beta \to \infty} \Pi^\beta(x) = |M|^{-1}\chi_M(x).$$

Proof. We recall $\Pi^\beta(x) = \exp(-\beta H(x))/\sum_{z \in X} \exp(-\beta H(z))$ and denote $H_0(x) = H(x) - \min_{z \in X} H(z)$. Since $H_0(x) = 0$ iff $x \in M$, we have

$$\Pi^\beta(x) = \frac{\exp(-\beta H_0(x))}{|M| + \sum_{z \notin M} \exp(-\beta H_0(z))} \overset{\beta \to \infty}{\longrightarrow} |M|^{-1}\chi_M(x). \qquad \square$$

4.2.4. *Simulated annealing.* Let $(\beta(n))_{n \geq 1}$ be a positive increasing sequence, a *cooling schedule*, and set

$$P_n = \Pi_{\{1\}}^{\beta(n)} \cdots \Pi_{\{\sigma\}}^{\beta(n)}, \qquad n \geq 1.$$

Consider the maximal local oscillation of H,

$$\Delta = \max_{s \in S} \max_{x_{S \setminus \{s\}} = y_{S \setminus \{s\}}} |H(x) - H(y)|.$$

If $\beta(n)$ increases to infinity such that $\beta(n) \leq \ln(n)/(|S|\Delta)$ eventually, then uniformly for all initial distributions ν,

$$\lim_{n \to \infty} \nu^T P_1 \cdots P_n = |M|^{-1}\chi_M.$$

4.2.5. *Metropolis sampling.* One samples from a Gibbs field $\Pi = Z^{-1}\exp(-H)$ in a two-step procedure, using an additional stochastic matrix G on X. The proposal step: Given x, sample y from $G(x, \cdot)$. The acceptance step: The proposal y is accepted with probability $\exp(-(H(y) - H(x))^+)$. In particular, y is accepted with probability one, if $H(y) \leq H(x)$. This results in the stochastic matrix

$$\pi(x, y) = \begin{cases} G(x, y)\exp(-(H(y) - H(x))^+), & x \neq y, \\ 1 - \sum_{z \neq x} \pi(x, z), & x = y. \end{cases}$$

The Metropolis sampler is capable of escaping local minima.

4.2.6. *Single flip algorithms.* Consider $X = X_*^S$ and set

$$G(x,y) = \begin{cases} \left(|S|(|X_*|-1)\right)^{-1}, & x_s \neq y_s \text{ for precisely one } s \in S, \\ 0, & \text{otherwise.} \end{cases}$$

This means to pick one site s at uniformly at random and then y_s uniformly at random. We have $\sum_{y \in X} G(x,y) = 1$ for all $x \in X$.

4.2.7. *Reversibility.* If G is a symmetric matrix, then the Metropolis sampler π is reversible with respect to Π.

Proof. We check detailed balance $\Pi(x)\pi(x,y) = \Pi(y)\pi(y,x)$ for all $x, y \in X$ by observing that

$$\exp(-H(x))\exp(-(H(y)-H(x))^+) = \exp(-H(y))\exp(-(H(x)-H(y))^+),$$

since $a + (b-a)^+ = a + \frac{1}{2}((b-a)+|b-a|) = b + \frac{1}{2}((a-b)+|a-b|) = b + (a-b)^+$ for all $a, b \in \mathbb{R}$. □

4.2.8. *Recommended reading.* [W, §5 & §10], [AKL]

4.3. **More Metropolis Algorithms (22.12.)**

4.3.1. *Metropolis sampler.* The Metropolis sampler of a Gibbs field Π with energy function H is defined by the transition matrix

$$\pi(x,y) = \begin{cases} G(x,y)\exp(-(H(y)-H(x))^+), & x \neq y, \\ 1 - \sum_{z\neq x} \pi(x,z), & x = y, \end{cases}$$

where G is the stochastic matrix for the proposal step. If G is symmetric, then π is reversible with respect to Π.

4.3.2. *Irreducibility.* A stochastic matrix A on X is called *irreducible*, if for all $x, y \in X$ there exist $x = z_0, z_1, \ldots, z_N = y \in X$ with $A(z_{j-1}, z_j) > 0$ for all $j = 1, \ldots, N$. The single flip proposal matrix G ("site and value uniformly at random") is irreducible.

4.3.3. *Convergence.* If G is symmetric and irreducible, then $c(\pi) < 1$ and

$$\lim_{n\to\infty} \nu^T \pi^n = \Pi^T$$

uniformly for all initial distributions ν.

4.3.4. *The eggbox.* Consider on $X = \{1, \ldots, m\}^2$ the energy

$$H(x) = \cos(\tfrac{\pi}{10}x_1)\cos(\tfrac{\pi}{10}x_2), \qquad x \in X,$$

and the proposal matrix $G(x,\cdot)$ that chooses for $x \in X$ one of its four neighbours uniformly at random.

```
h = @(x) cos(pi*x(1)/10).*cos(pi*x(2,:)/10); ...
m = 100; xx = [1:m]; [X1,X2] = meshgrid(xx); ...
H = cos(pi*X1/10).*cos(pi*X2/10); contour(X1,X2,H), ...
x = randi(50,2,1); hold on, plot(x(1),x(2),'r*'), drawnow, ...
for k = 1:1e+3, y = x;  u = randi(4); ...
if (u==1), y(1) = x(1) - 1; elseif (u==2), y(1) = x(1) + 1; ...
elseif (u==3), y(2) = x(2) - 1; else, y(2) = x(2) + 1; end, ...
if (y(1)==0), y(1)=m; end, if (y(1)==m+1), y(1)=1; end,...
if (y(2)==0), y(2)=m; end, if (y(2)==m+1), y(2)=1; end,...
u = rand; if (u<=exp(h(x)-h(y))), x = y; end, ...
plot(x(1),x(2),'r*'), drawnow, end,
```

4.3.5. *Threshold random search.* Let x be given and y be the proposal. Let u be uniformly distributed in $[0,1]$. The Metropolis sampler accepts y, if

$$u \leq \exp(H(x) - H(y)), \quad \text{that is,} \quad H(y) - H(x) \leq -\ln(u).$$

With $-\ln(u)$ as a nonnegative random threshold, one can view the Metropolis sampler as a threshold random search algorithm.

4.3.6. *Positive random fields.* For a positive random field Π the Metropolis sampler with proposal matrix G is given by

$$\pi(x,y) = \begin{cases} G(x,y)\min\{\Pi(y)/\Pi(x), 1\}, & x \neq y, \\ 1 - \sum_{z\neq x} \pi(x,z), & x = y. \end{cases}$$

4.3.7. *Metropolis–Hastings method.* Consider $A : X \times X \to [0,1]$ and set

$$\pi(x,y) = \begin{cases} G(x,y)A(x,y), & x \neq y, \\ 1 - \sum_{z \neq x} \pi(x,z), & x = y. \end{cases}$$

If $G(x,y)$ and $G(y,x)$ are either both positive or zero, one might choose the acceptance probability

$$A(x,y) = \min\left\{ \frac{\Pi(y)G(y,x)}{\Pi(x)G(x,y)}, 1 \right\},$$

such that π is reversible with respect to Π.

4.3.8. *Gibbs sampling revisited.* The Gibbs sampler with uniformly random visiting scheme is a Metropolis–Hastings sampler with acceptance probability equal to one and proposal matrix

$$G(x,y) = \frac{1}{|S|} \sum_{s \in S} \Pi(y_s \mid x_{S\setminus\{s\}}) \chi_{\{x_{S\setminus\{s\}} = y_{S\setminus\{s\}}\}}.$$

Indeed, for $x \neq y$, at most one summand of $G(x,y)$ is non-zero, and $G(x,y)$ and $G(y,x)$ are either both positive or zero. In the positive case, we have $x_{S\setminus\{s\}} = y_{S\setminus\{s\}}$ and $\Pi(x)G(x,y) = |S|^{-1}\Pi(x)\Pi(y_s \mid x_{S\setminus\{s\}}) = \Pi(y)G(y,x)$, that is, $A(x,y) = 1$.

4.3.9. *Recommended reading.* [W, §10], [MNR, 6.3]

4.4. **Convergence rates (12.01.)**

4.4.1. *Convergence rate in terms of the ergodicity coefficient.* Let μ be stationary with respect to a stochastic matrix P. Then, for all initial distributions ν,

$$\left\|\nu^T P^n - \mu^T\right\|_1 = \left\|(\nu-\mu)^T P^n\right\|_1 \le c(P^n)\|\nu-\mu\|_1 \le 2\,c(P)^n.$$

Proof. We check, that P leaves $\{w \in \ell^1 : \|w\|_1 \le 1, \sum_j w_j = 0\}$ invariant:

$$\|w^T P\|_1 = \sum_j |\sum_i w_i p_{ij}| \le \sum_i |w_i| = \|w\|_1,$$

and similarly $\sum_j (w^T P)_j = \sum_{i,j} w_i p_{ij} = \sum_i w_i$. □

4.4.2. *The spectral gap.* Let the finite stochastic matrix P be reversible with respect to μ. Then, $\langle x, Py\rangle_\mu = \langle Px, y\rangle_\mu$ for all $x, y \in \ell^2$. Moreover, $\sigma(P) \subset [-1,1]$ and $1 \in \sigma(P)$ as well as

$$\lambda_* := \max\{|\lambda| : \lambda \in \sigma(P), \lambda \ne 1\} \le c(P).$$

Proof. If $\lambda \ne 1$ and $w^T P = \lambda w^T$, then $\sum_i w_i p_{ij} = \lambda w_j$ for all j and

$$\sum_i w_i = \sum_{i,j} w_j p_{ij} = \lambda \sum_j w_j,$$

that is, $\sum_j w_j = 0$. Therefore, $|\lambda| = \|w^T P\|_1 / \|w\|_1 \le c(P)$. □

4.4.3. *Rayleigh–Ritz characterization.* Let the finite stochastic matrix P be reversible with respect to μ. If $\lambda = 1$ is a simple eigenvalue of P, then

$$1 - \lambda_*^2 = \min\left\{ \frac{\langle (1-P^2)f, f\rangle_\mu}{\mathbb{V}_\mu(f)} \,:\, \mathbb{E}_\mu(f) = 0 \right\}.$$

Proof. The minimization is over the subspace μ-orthogonal to the constants, that is, for vectors f with $0 = \langle f, \mathbf{1}\rangle_\mu = \mathbb{E}_\mu(f)$. In this case, $\mathbb{V}_\mu(f) = \langle f, f\rangle_\mu$. □

4.4.4. *Convergence rate in terms of the spectral gap.* Let P be a finite positive stochastic matrix that is reversible with respect to μ. Then, μ is strictly positive, and we have for all initial distributions ν,

$$\left\|\nu^T P^n - \mu^T\right\|_1 \le \sigma_\mu(\tfrac{\nu}{\mu})\,\lambda_*^n.$$

Proof. We denote $\nu_n = (\nu^T P^n)^T$ and $\rho_n = \nu_n/\mu$. We prove $\mu > 0$ in the exercises. Convexity of the square function implies

$$\|\nu_n - \mu\|_1^2 = \left(\sum_i \frac{|(\nu_n)_i - \mu_i|}{\mu_i}\mu_i\right)^2 \le \sum_i \frac{((\nu_n)_i - \mu_i)^2}{\mu_i^2}\mu_i.$$

Moreover, $(\rho_n)_i - \mathbb{E}_\mu(\rho_n) = \frac{(\nu_n)_i}{\mu_i} - \sum_j \frac{(\nu_n)_j}{\mu_j}\mu_j = \frac{(\nu_n)_i}{\mu_i} - 1 = \frac{(\nu_n)_i - \mu_i}{\mu_i}$ so that

$$\left\|\nu_n - \mu\right\|_1 \le \sigma_\mu(\rho_n).$$

By reversibility,

$$(P\rho_n)_i = \sum_j p_{ij}\tfrac{(\nu_n)_j}{\mu_j} = \sum_j p_{ji}\tfrac{(\nu_n)_j}{\mu_i} = (\rho_{n+1})_i,$$

so that

$$\mathbb{V}_\mu(\rho_{n+1}) = \mathbb{V}_\mu(P\rho_n) \overset{\text{Ex.}}{=} \mathbb{V}_\mu(\rho_n) - \langle(\mathrm{Id} - P^2)\rho_n, \rho_n\rangle_\mu.$$

The Rayleigh–Ritz characterization implies $(1-\lambda_*^2)\mathbb{V}_\mu(\rho_n) \geq \langle (1-P^2)\rho_n, \rho_n \rangle_\mu$ and

$$\mathbb{V}_\mu(\rho_{n+1}) \leq \big(1 - (1-\lambda_*^2)\big)\, \mathbb{V}_\mu(\rho_n) = \lambda_*^2\, \mathbb{V}_\mu(\rho_n) \leq \lambda_*^{2n}\, \mathbb{V}_\mu(\rho_0). \qquad \square$$

4.4.5. *Recommended reading.* [B, §6], [W, §11]

5. Quasi–Monte Carlo

5.1. **Convergence test and quasi–Monte Carlo integration (19.01.)**

5.1.1. *Gelman–Rubin test.* A simplified Gelman–Rubin test for estimating

$$\mathbb{E}_\mu(f) = \int f d\mu$$

proceeds as follows:

(1) Simulate m independent Markov chains of length $2n$ with stationary distribution μ and discard the first n draws of each chain. One obtains f_{ij} with $i = 1, \ldots, m$ and $j = 1, \ldots, n$.
(2) Calculate the chain means $\overline{f_i}$ and the estimator $\overline{f}$ of the target mean,

$$\overline{f_i} = \frac{1}{n} \sum_{j=1}^{n} f_{ij}, \qquad \overline{f} = \frac{1}{m} \sum_{i=1}^{m} \overline{f_i}.$$

(3) Calculate the average of the within-chain variance,

$$W = \frac{1}{m} \sum_{i=1}^{m} s_i^2 \quad \text{with} \quad s_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} \left(f_{ij} - \overline{f_i}\right)^2.$$

(4) Calculate the between-chain variance,

$$\frac{1}{n} B = \frac{1}{m-1} \sum_{i=1}^{m} \left(\overline{f_i} - \overline{f}\right)^2.$$

(5) Estimate the target variance by $\hat{V}_n = (1 - \frac{1}{n})W + \frac{1}{n}B$.
(6) Check, if the ratio $\hat{R} = \sqrt{\hat{V}_n / W}$ is sufficiently close to one.

One might work with $m = 5$ or $m = 10$ independent chains and increase the chain length $2n$ until $|\hat{R} - 1|$ is below 0.2 or 0.1.

5.1.2. *Unbiased estimates.* If the chains are stationary, then $\overline{f}$ and $\hat{V}_n$ are unbiased estimators for $\mathbb{E}_\mu(f)$ and $\mathbb{V}_\mu(f)$, respectively.

Proof. We calculate $\mathbb{E}_\mu(\overline{f}) = \mathbb{E}_\mu(\overline{f_i}) = \mathbb{E}_\mu(f)$ and for all $a \in \mathbb{R}$,

$$\begin{aligned} \hat{V}_n &= \frac{1}{mn} \sum_{i,j} \left(f_{ij} - \overline{f_i}\right)^2 + \frac{1}{m-1} \sum_i \left(\overline{f_i} - \overline{f}\right)^2 \\ &= \frac{1}{mn} \sum_{ij} (f_{ij} - a)^2 + \frac{1}{m(m-1)} \sum_i \left(\overline{f_i} - a\right)^2 - \frac{m}{m-1} \left(\overline{f} - a\right)^2. \end{aligned}$$

With $a = \mathbb{E}_\mu(f)$ we have

$$\frac{m}{m-1} \mathbb{E}_\mu \left((\overline{f} - a)^2\right) = \frac{1}{m(m-1)} \sum_i \mathbb{E}_\mu((\overline{f_i} - a)^2),$$

so that $\mathbb{E}_\mu(\hat{V}_n) = \frac{1}{mn} \sum_{ij} \mathbb{E}_\mu((f_{ij} - a)^2) = \mathbb{V}_\mu(f)$. □

5.1.3. *Discrepancy.* Let $f \in \mathcal{S}(\mathbb{R}^d)$ and μ a probability distribution on $\mathbb{R}^d$ such that $f \in L^1(\mu)$. Then, for all $z_1, \ldots, z_n \in \mathbb{R}^d$,

$$\frac{1}{n}\sum_{j=1}^n f(z_j) - \int_{\mathbb{R}^d} f(x)\mu(dx) = (-1)^d \int_{\mathbb{R}^d} \partial^{1:d} f(x) \mathcal{D}_\mu(z_1, \ldots, z_n; x) dx,$$

where

$$\mathcal{D}_\mu(z_1, \ldots, z_n; x) = \frac{1}{n}\sum_{j=1}^n \chi_{]-\infty,x]}(z_j) - \mu(]-\infty, x]), \qquad x \in \mathbb{R}^d,$$

denotes the *discrepancy* function with respect to the distribution μ.

Proof. We write

$$f(x) = -\int_{x_1}^\infty \partial_1 f(y_1, x_2, \ldots, x_d) dy_1 = (-1)^d \int_{[x,\infty[} \partial^{1:d} f(y) dy,$$

so that

$$\frac{1}{n}\sum_{j=1} f(z_j) = (-1)^d \int_{\mathbb{R}^d} \partial^{1:d} f(y) \frac{1}{n}\sum_{j=1}^n \chi_{]-\infty,y]}(z_j) dy$$

and

$$\int_{\mathbb{R}^d} f(x)\mu(dx) = (-1)^d \int_{\mathbb{R}^d} \partial^{1:d} f(y) \mu(]-\infty, y]) dy. \qquad \square$$

5.1.4. *Koksma-Hlawka inequality.* For all $z_1, \ldots, z_n \in \mathbb{R}^d$,

$$\left| \frac{1}{n}\sum_{j=1}^n f(z_j) - \int_{\mathbb{R}^d} f d\mu \right| \le \int_{\mathbb{R}^d} \left| \partial^{1:d} f(x) \right| dx \sup_{x \in \mathbb{R}^d} \left| \mathcal{D}_\mu(z_1, \ldots, z_n; x) \right|.$$

5.1.5. *Integration on the unit cube.* The Koksma–Hlawka inequality also holds for equiweighted quadrature on the unit cube $[0,1]^d$, however, the variation of the integrand f has to be augmented by boundary values of f and its derivatives.

5.1.6. *Low discrepancy sets.* For the uniform distirbution μ on the unit cube $[0,1]^d$, several point sets $\{z_1, \ldots, z_n\} \subset [0,1]^d$ with

$$\sup_{x \in \mathbb{R}^d} |\mathcal{D}_\mu(z_1, \ldots, z_n; x)| = O((\log n)^{d-1}/n)$$

are known, e.g. Halton or Sobol sets.

```
n = 100; m = rand(n,2); p = haltonset(2); q = net(p,n); ...
plot(m(:,1),m(:,2),'bo',q(:,1),q(:,2),'go'),...
N = [1:1e+3]; for n=N, m = rand(n,2); q = net(p,n); ...
mm(n) = mean(m(:,1)); qq(n) = mean(q(:,1)); end,...
semilogy(N,abs(mm-0.5),'b',N, abs(qq-0.5),'g'), hold on, ...
semilogy(N, 1./sqrt(N),'r',N,1./N, 'r')
```

5.1.7. *Recommended reading.* [GR], [NW10, 9]

6. Exam (02.02.)

References

[AFB] S. El Adlouni, A. Favre, B. Bobée, Comparison of methodologies to assess the convergence of Markov chain Monte Carlo methods, Comp. Stat. Data An. **50**, 2685–2701, 2006.

[AKL] E. Aarts, J. Korst, P. Laarhoven, Simulated Annealing, Local Search in Combinatorial Optimization, E. Aarts and J. Lenstra editors, Wiley 1997, 91–120.

[AG] S. Asmussen, P. Glynn, Stochastic Simulation: Algorithms and Analysis, Springer-Verlag, 2007.

[B] P. Brémaud, Markov Chains, Gibbs Fields, Monte Carlo Simulation and Queues, Springer-Verlag, 1999.

[De] L. Devroye, Non-Uniform Random Variate Generation, Springer-Verlag, 1986.

[Di] P. Diaconis, The Markov Chain Monte Carlo Revolution, Bull. Amer. Math. Soc. **46**, 179–205, 2009.

[GR] A. Gelman, D. Rubin, Inference from iterative simulation using multiple sequences, Stat. Sci. **7**, 457–511, 1992.

[H] O. Häggström, Finite Markov Chains and Algorithmic Applications, Cambridge University Press, 2002.

[L] G. Lawler, Introduction to Stochastic Processes, 2n edition, Taylor & Francis, 2006.

[MNR] T. Müller-Gronbach, E. Novak, K. Ritter, Monte Carlo-Algorithmen, Springer-Verlag, 2012.

[MT] S. Meyn, R. Tweedie, Markov Chains and Stochastic Stability, Springer-Verla,g 1993.

[NW08] E. Novak, H. Wozniakowski, Tractability of Multivariate Problems, Volume I. Linear Information, EMS 2008.

[NW10] E. Novak, H. Wozniakowski, Tractability of Multivariate Problems, Volume II. Standard Information for Functionals, EMS 2010.

[RC] C. Robert, G. Casella, Monte Carlo Statistical Methods, Seond Edition, Springer-Verlag, 2004.

[S] J. Sanz-Serna, Markov Chain Monte Carlo and numerical differential equations, Current Challenges in Stability Issues for Numerical Differential Equations, L. Dieci and N. Guglielmi editors, Lect. Notes in Math., Vol. 2082, Springer, 2014, 39–88.

[SM] R. Shonkwiler, F. Mendivil, Explorations in Monte Carlo Methods, Springer-Verlag, 2009.

[W] G. Winkler, Image Analysis, Random Fields, and Markov Chain Monte Carlo Methods, 2nd edition, Springer-Verlag, 2003.

[ZZ] R. Zaharopol, G. Zbaganu, Dobrushin Coefficients of Ergodicity and Asymptotically Stable L^1-Contractions, Jour. Theo. Prob. **12**, 885–902, 1999.