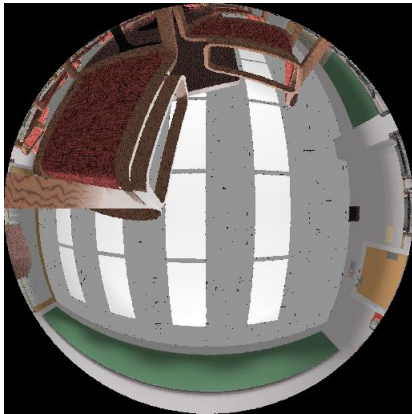

Počítačová grafika III – Monte Carlo integrování

Jaroslav Křivánek, MFF UK

Jaroslav.Krivanek@mff.cuni.cz

Rendering = Integrovaní funkcí

$$L_r(\mathbf{x}, \omega_o) = \int_{H(\mathbf{x})} L_i(\mathbf{x}, \omega_i) \cdot f_r(\mathbf{x}, \omega_i \rightarrow \omega_o) \cdot \cos \theta_i \, d\omega_i$$



■ Problémy

- ❑ Nespojitosť integradu (viditelnost)
- ❑ Téměř libovolné hodnoty integrandu (distribuce světla, BRDF)
- ❑ Složitá geometrie



↖
Příchozí radiance
 $L_i(\mathbf{x}, \omega_i)$ pro jeden
bod na podlaze.

Historie Monte Carlo (MC)

- Vývoj atomové bomby, Los Alamos 1940, John von Neumann, Stanislav Ulam, Metropolis
- Rozvoj a aplikace metod od roku 1949

Metoda Monte Carlo

- Simuluje se mnoho případů daného děje, například:
 - Neurony – vznik, zánik, srážky s atomy vodíku
 - Úlohy hromadné obsluhy – chování počítačových sítí, dopravní situace
 - Sociologické a ekonomické modely – demografie, vývoj inflace, pojišťovnictví atd.

Odbočka – Kvadraturní vzorce pro numerické integrování

- Obecný předpis v 1D:

$$\hat{I} = \sum_{i=1}^n w_i f(x_i)$$

f	integrand (tj. integrovaná funkce)
n	řád kvadratury (tj. počet vzorků integrandu)
x_i	uzlové body (tj. umístění vzorků v oboru integrálu)
$f(x_i)$	vzorky integrandu
w_i	váhy

Odbočka – Kvadraturní vzorce pro numerické integrování

- Kvadraturní pravidla se liší volbou uzlových bodů x_i a váhami w_i
 - Obdélníková metoda, Rovnoběžníková metoda, Simpsonova metoda, Gaussovská kvadratura, ...
- Vzorčky na integračním oboru (tj. uzlové body) jsou rozmístěny deterministicky
 - Jednoznačně určeny kvadraturním pravidlem

Kvadrurní vzorce pro více dimenzí

- Obecný předpis pro integrování fcí více proměnných:

$$\hat{I} = \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_s=1}^n w_{i_1} w_{i_2} \dots w_{i_s} f(x_{i_1}, x_{i_2}, \dots, x_{i_s})$$

- Rychlost konvergence pro s -dimenzionální integrál je $O(N^{-1/s})$
 - Např. pro dvojnásobné zpřesnění odhadu 3-rozměrného integrálu musíme zvýšit počet vzorků $2^3 = 8$ krát
- Nepoužitelné pro vysokodimenzionální integrály
 - **Dimenzionální exploze**

Kvadrurní vzorce pro více dimenzí

■ Kvadrurní vzorce

- V 1D lepší přesnost než Monte Carlo
- Ve 2D srovnatelné s MC
- Od 3D bude MC téměř vždy lepší

■ Kvadrurní metody NEJSOU metody Monte Carlo!

Monte Carlo integrování

- Vzorky jsou rozmístěny náhodně (nebo pseudonáhodně)
- Konvergence: $O(N^{-1/2})$
 - **Konvergence nezávisí na dimenzionalitě**
 - Rychlejší než klasické kvadrurní vzorce pro 3 a více dimenzí
- Speciální metody pro rozmístění vzorků
 - Quasi-Monte Carlo, Randomized quasi-Monte Carlo
 - Ještě rychlejší konvergence než MC

Monte Carlo integrování – shrnutí

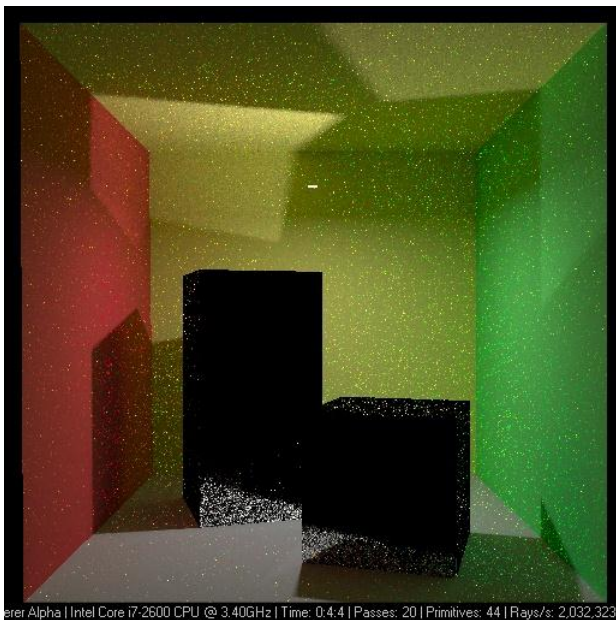
■ Výhody

- ❑ Jednoduchá implementace
- ❑ Robustní řešení pro různé tvary domén a integrantů
- ❑ Efektivní pro vícerozměrné integrály

■ Nevýhody

- ❑ Relativně pomalá konvergence – zmenšení statistické chyby o polovinu vyžaduje zvětšit počet vzorků čtyřikrát
- ❑ Pro syntézu obrazu: obrázek obsahuje šum

Šum v obrázcích



Náhodné veličiny

Náhodná veličina

- X ... náhodná veličina
- X nabývá různých hodnot s různou pravděpodobností
 - $X \sim p(x)$
 - Rozložení pravděpodobnosti

Diskrétní náhodná veličina

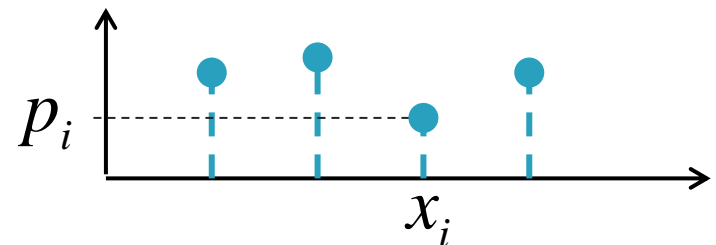
- Konečná množina hodnot x_i
- S pravděpodobnostmi p_i

$$p_i \equiv \Pr(X = x_i) \geq 0 \quad \sum_{i=1}^n p_i = 1$$

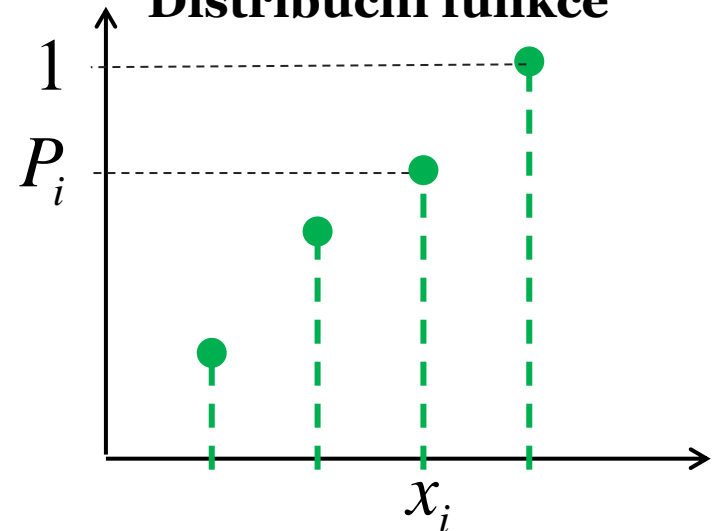
- **Distribuční funkce**
(cumulative distribution function)

$$P_i \equiv \Pr(X \leq x_i) = \sum_{j=1}^i p_j \quad P_n = 1$$

Pravděpodobnostní funkce (probability mass function)



Distribuční funkce



Spojité náhodná veličina

- **Hustota pravděpodobnosti** $p(x)$
(probability density function, **pdf**)

$$\Pr(X \in D) = \int_D p(x) dx$$

- V 1d:

$$\Pr(a < X \leq b) = \int_a^b p(t) dt$$

Spojité náhodná veličina

- **Distribuční funkce** $P(x)$
(cumulative distribution function, **cdf**)
V 1d:

$$P(x) \equiv \Pr(X \leq x) = \int_{-\infty}^x p(t) dt$$

$$\Pr(X = a) = \int_a^a p(t) dt = 0!$$

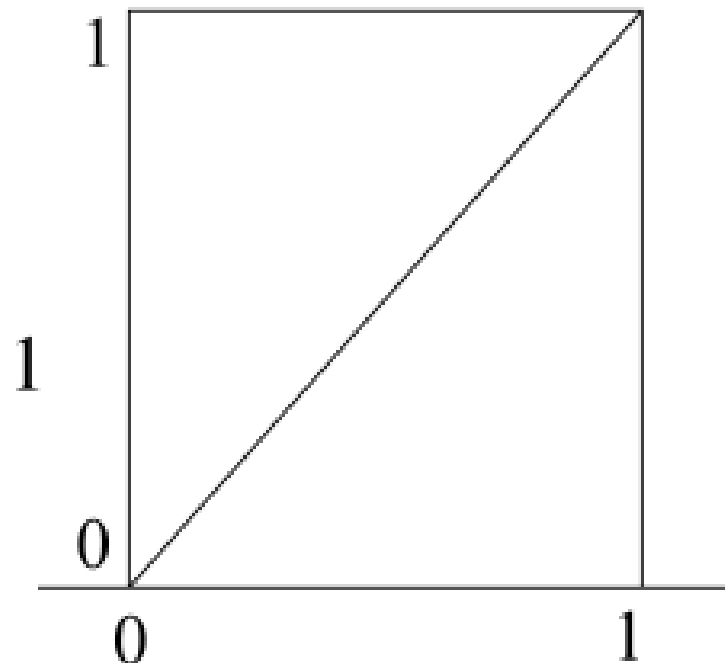
Spojité náhodná veličina

Př. Rovnoměrné rozdělení (uniform distribution)

Hustota
pravděpodobnosti
(pdf)



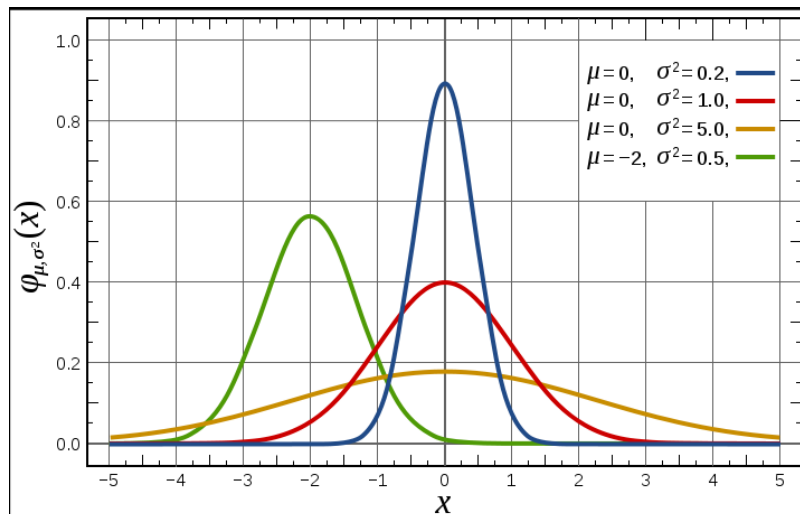
Distribuční
funkce (cdf)



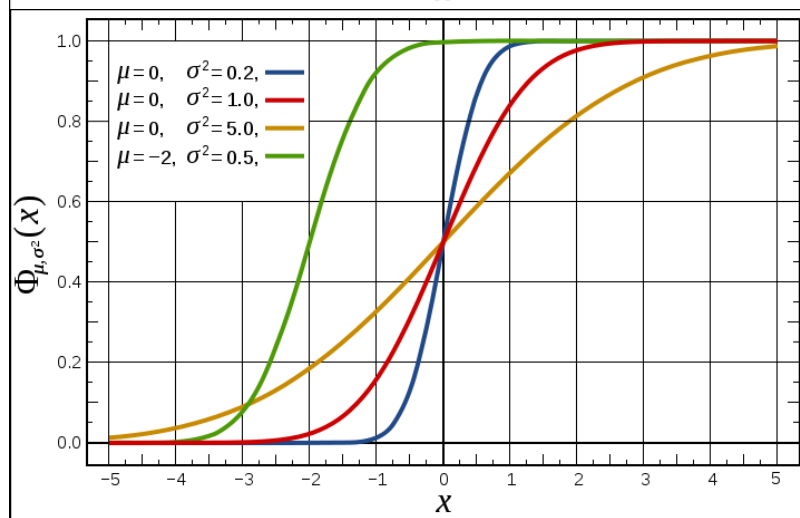
Spojité náhodná veličina

Gaussové (normální) rozdělení

Hustota
pravděpodobnosti
(pdf)



Distribuční
funkce (cdf)



Zdroj: wikipedia

Střední hodnota a rozptyl

- **Střední hodnota** (očekávaná hodnota, expected value)

$$E[X] = \int_D \mathbf{x} p(\mathbf{x}) d\mathbf{x}$$

- **Rozptyl** (variance)

$$\begin{aligned} V[X] &= E[(X - E[X])^2] \\ &= E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - E[X]^2 \end{aligned}$$

- Vlastnosti

$$V\left[\sum_i X_i\right] = \sum_i V[X_i] \quad (\text{pokud jsou } X_i \text{ nezávislé})$$

$$V[aX] = a^2 V[X]$$

Transformace náhodné veličiny

$$Y = f(X)$$

- Y je náhodná veličina
- Střední hodnota Y

$$E[Y] = \int_D f(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}$$

Monte Carlo integrování

Primární estimátor určitého integrálu

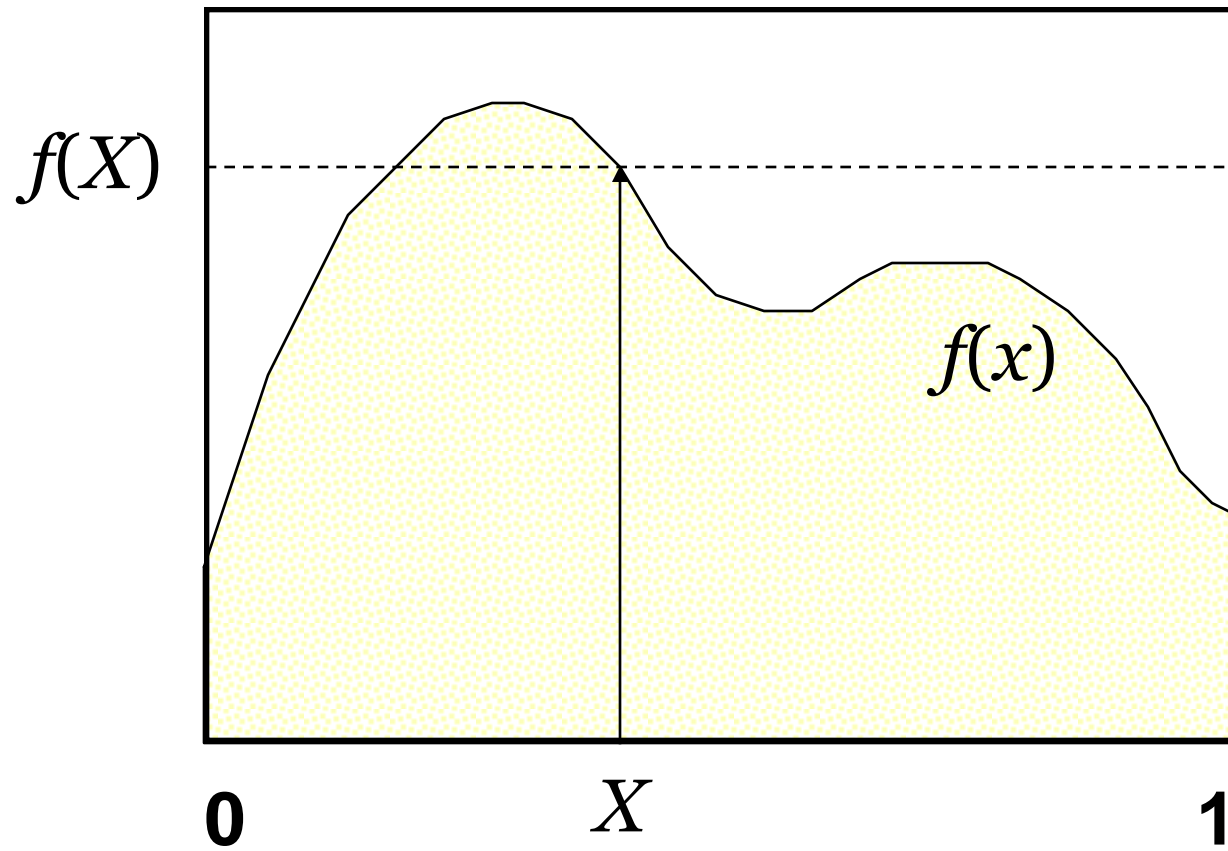
Odhadovaný integrál:

$$I = \int_{\Omega} f(x) dx$$

Je-li X náhodná veličina s distribucí $p(x)$, pak $Y = f(X)/p(X)$ je tzv. **primární estimátor** integrálu:

$$F_{\text{prim}} = Y = \frac{f(X)}{p(X)}$$

Primární estimátor určitého integrálu

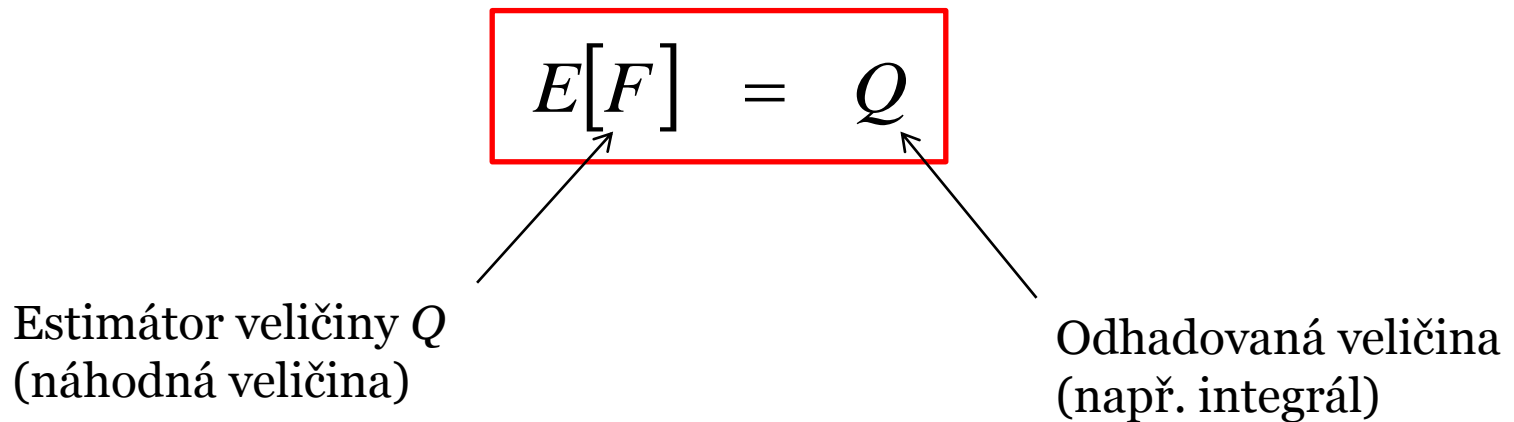


Estimátor a odhad

- **Estimátor je náhodná veličina**
 - Vznikla transformací jiné náhodné veličiny
- Její **realizace** (hodnota) je konkrétní **odhad** (**estimate**)

Nestrannost obecného estimátoru

- **Nestrannost estimátoru (obecně):**
 - „V průměru“ estimátor dává správnou hodnotu odhadované veličiny (bez systematické chyby)



Nestrannost

Náš estimátor F_{prim} je nestranným (**unbiased**) odhadem I

$$\begin{aligned} E[F_{\text{prim}}] &= E[Y] \\ &= \int_{\Omega} \frac{f(x)}{p(x)} p(x) dx \\ &= I \end{aligned}$$

Rozptyl primárního estimátoru

Měřítkem kvality odhadu je jeho **rozptyl** (nebo standardní odchylka):

$$\underline{V[F_{\text{prim}}] = \sigma_{\text{prim}}^2 = E[Y^2] - E[Y]^2 = \int_{\Omega} \frac{f(x)^2}{p(x)} dx - I^2}$$

(pro nestranný odhad)

Při výpočtu **jediného vzorku** je rozptyl výsledku příliš velký!

Sekundární estimátor integrálu

- N nezávislých náhodných veličin, $Y_i = f(X_i) / p(X_i)$

$$\begin{aligned} F_N &= \frac{1}{N} \sum_{i=1}^N Y_i \\ &= \frac{1}{N} \sum_{i=1}^N \frac{f(X_i)}{p(X_i)} \end{aligned}$$

- Sekundární estimátor je **nestranný**

Rozptyl sekundárního estimátoru

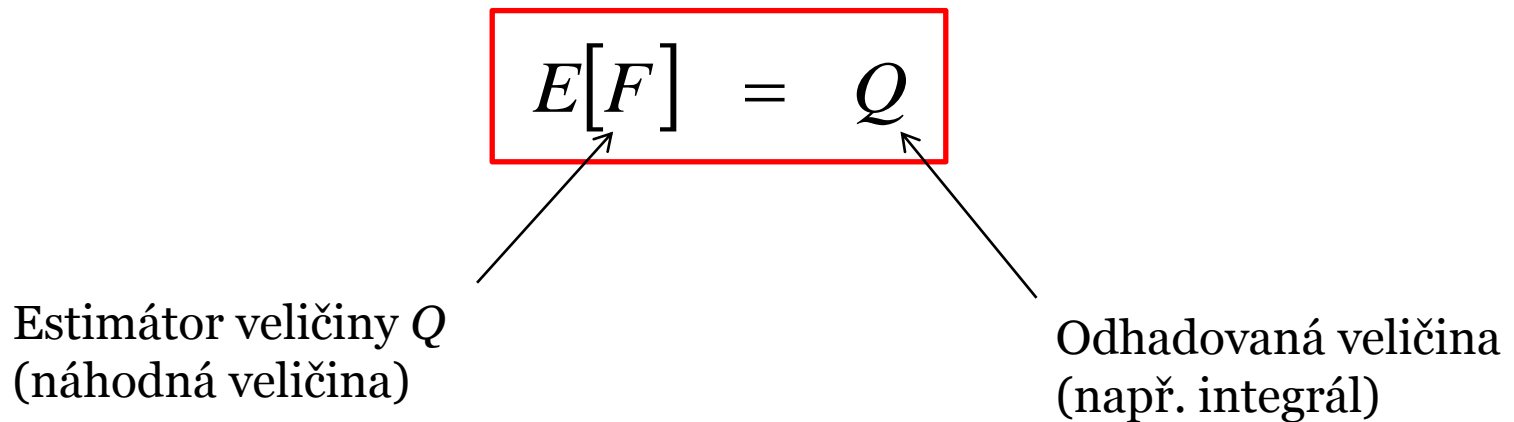
$$\begin{aligned}V[F_N] &= V\left[\frac{1}{N} \sum_{i=1}^N Y_i\right] \\&= \frac{1}{N^2} \cdot N \cdot V[Y_i] \\&= \frac{1}{N} V[Y_i] \\&= \frac{1}{N} V[F_{\text{prim}}]\end{aligned}$$

... std. chyba je \sqrt{N} -krát menší!
(konvergence $1/\sqrt{N}$)

Vlastnosti estimátorů

Nestrannost obecného estimátoru

- **Nestrannost estimátoru (obecně):**
 - „V průměru“ estimátor dává správnou hodnotu odhadované veličiny (bez systematické chyby)



Výchylka (bias) obecného estimátoru

- Pokud

$$E[F] \neq Q$$

- pak estimátor není nestranný (je vychýlený, „**biased**“).
- Systematická chyba, **bias**

$$\beta = Q - E[F]$$

Konzistence (obecného estimátoru)

- Uvažujme sekundární estimátor (N vzorků):

$$F_N = F_N(X_1, X_2, \dots, X_N)$$

- Estimátor F_N je **konzistentní** pokud

$$Pr \left\{ \lim_{N \rightarrow \infty} F_N = Q \right\} = 1$$

tj. pokud **chyba** $F_N - Q$ jde k nule s pravděpodobností 1.

Konzistence (obecného estimátoru)

- Postačující podmínka pro konzistenci estimátoru:

$$\lim_{N \rightarrow \infty} \beta[F_N] = \lim_{N \rightarrow \infty} V[F_N] = 0$$

↑
bias

(tj. ne každý nestranný estimátor je konzistentní)

Zobrazovací algoritmy

- **Nestranné (unbiased)**

- Sledování cest (path tracing)
- Obousměrné sledování cest (bidirectional path tracing)
- Metropolis light transport

- **Konzistentní (consistent)**

- Progresivní fotonové mapy (progressive photon mapping)

- **Nekonzistentní, vychýlené (biased)**

- Fotonové mapy (photon mapping)
- Irradiance / radiance caching

Střední kvadratická chyba (Mean Squared Error – MSE)

■ Definice

$$MSE[F] = E[(F - Q)^2]$$

■ Platí

$$MSE[F] = V[F] + \beta[F]^2$$

□ Důkaz

$$\begin{aligned} MSE[F] &= E[(F - Q)^2] \\ &= E[(F - E[F])^2] + 2E[F - E[F]](E[F] - Q) + (E[F] - Q)^2 \\ &= V[F] + \beta[F]^2, \end{aligned}$$

Střední kvadratická chyba (Mean Squared Error – MSE)

- Pokud F je nestranný, pak

$$MSE[F] = V[F]$$

tj. pro nestranný estimátor je snazší odhadnout chybu, protože rozptyl estimátoru lze odhadnout ze vzorků Y_i :

- Nestranný **estimátor rozptylu**

$$\hat{V}[F_N] = \frac{1}{N-1} \left\{ \left(\frac{1}{N} \sum_{i=1}^N Y_i^2 \right) - \left(\frac{1}{N} \sum_{i=1}^N Y_i \right)^2 \right\}$$

Účinnost estimátoru

- Pro nestranný estimátor je **účinnost** (eficience, angl. efficiency) dána vztahem:

$$\epsilon[F] = \frac{1}{V[F] T[F]}$$

rozptyl

čas výpočtu (počet operací, např. počet vržených paprsků)

Metody snížení rozptylu MC estimátorů

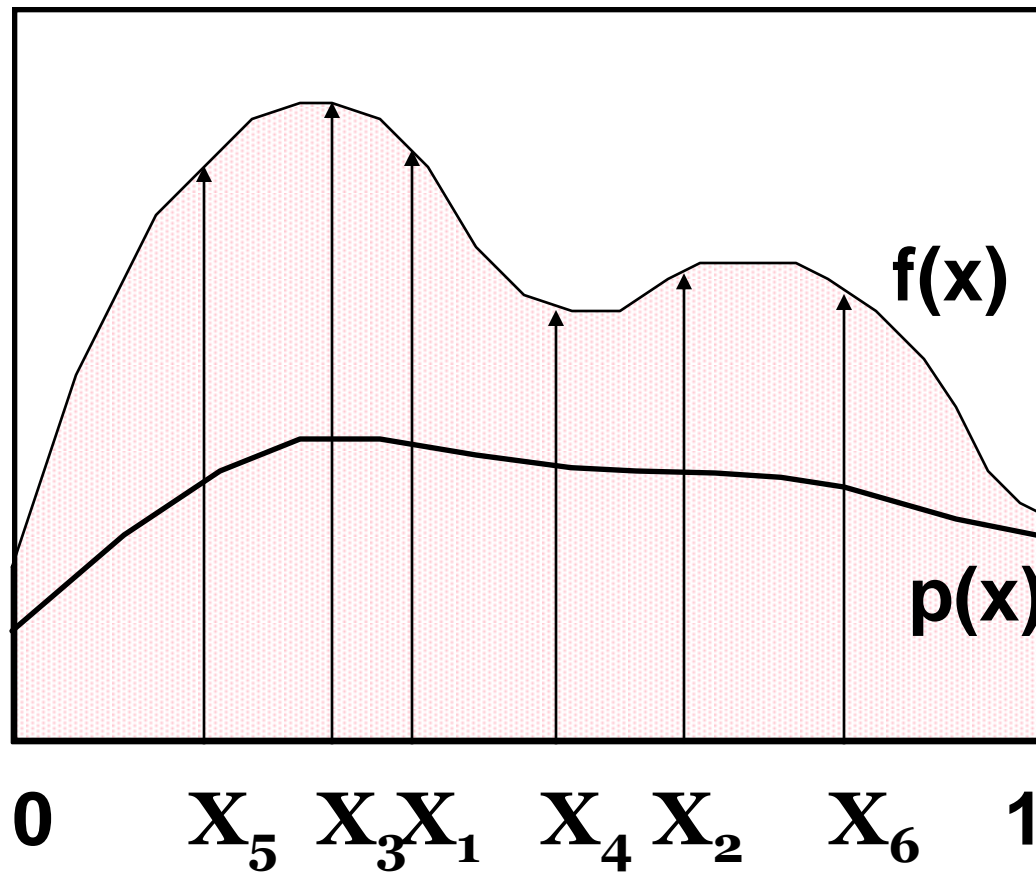
Metody snížení rozptylu

- Importance sampling (vzorkování podle důležitosti)
 - a) Podle BRDF (nejčastější)
 - b) Podle L_i (pokud známo: přímé osvětlení)
 - V syntéze obrazu je IS **nejčastěji používaná metoda**
- Řídící funkce (control variates)
- Lepší rozložení vzorků
 - Stratifikace
 - quasi-Monte Carlo (QMC)

Vzorkování podle důležitosti

- Některé části vzorkovaného intervalu jsou **důležitější**, protože zde má f větší hodnotu
 - Vzorky z těchto oblastí mají větší vliv na výsledek
- Vzorkování podle důležitosti (“**importance sampling**”) umisťuje vzorky přednostně do takových oblastí
 - Tj. $\text{pdf } p$ je „podobná“ integrandu
- **Menší rozptyl** při zachování nestrannosti

Vzorkování podle důležitosti



Řídící funkce

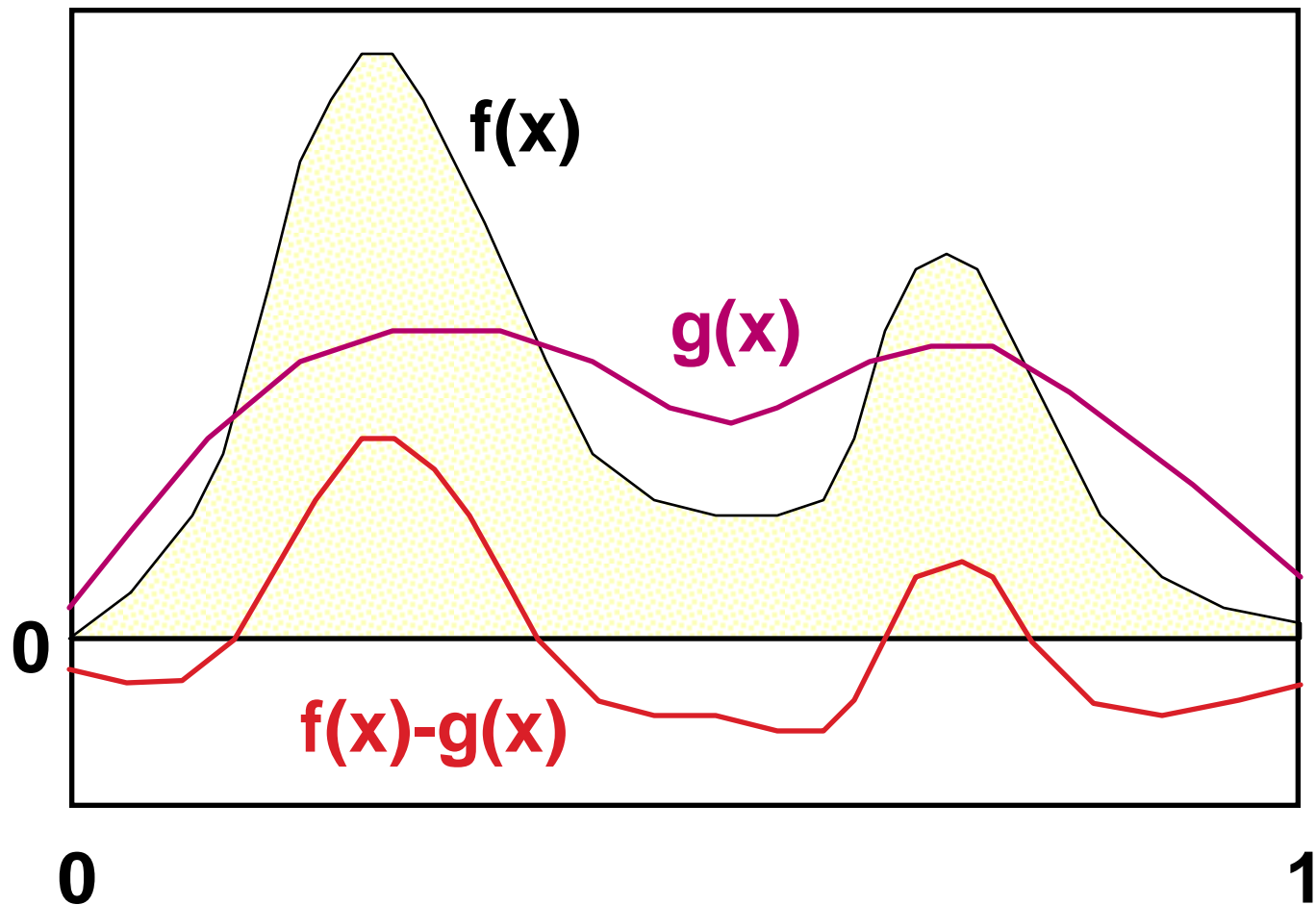
Funkce $g(\mathbf{x})$, která **aproximuje integrant** a dokážeme ji **analyticky integrovat**:

$$I = \int f(\mathbf{x}) d\mathbf{x} = \underbrace{\int [f(\mathbf{x}) - g(\mathbf{x})] d\mathbf{x}}_{\text{numerické integrování (MC) menší rozptyl než } f(\mathbf{x})} + \underbrace{\int g(\mathbf{x}) d\mathbf{x}}_{\text{umíme analyticky integrovat}}$$

numerické integrování (MC)
menší rozptyl než $f(\mathbf{x})$

umíme analyticky
integrovat

Transformace řídicí funkcí



Řídící funkce vs. Importance sampling

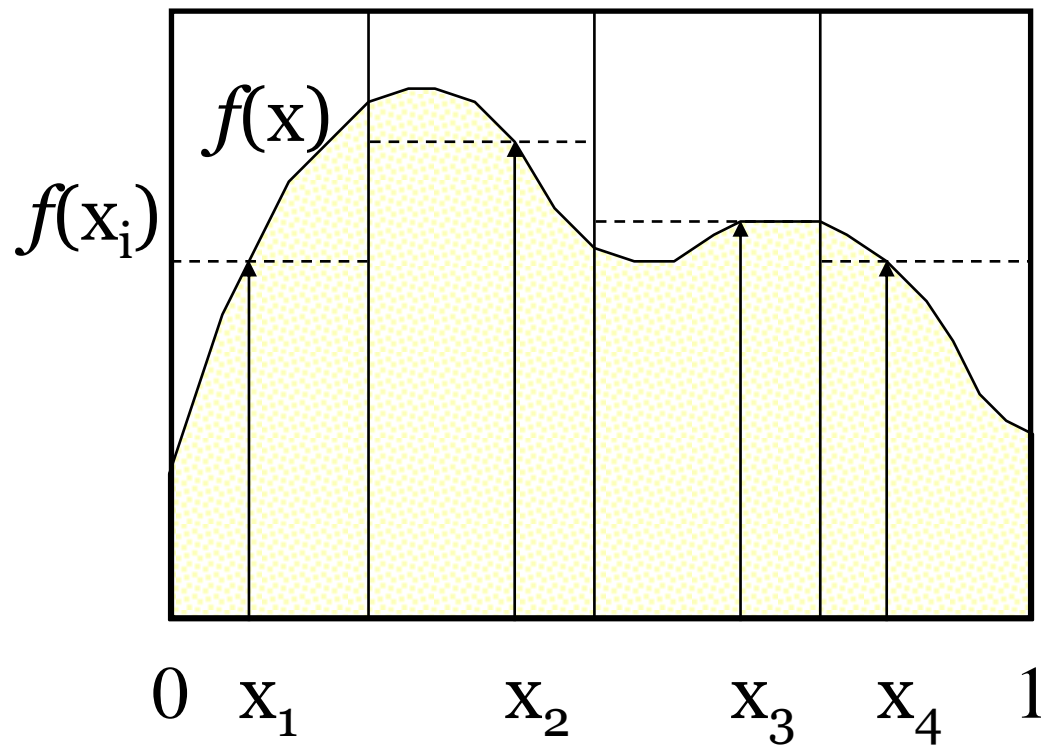
- Importance sampling
 - Lepší pokud se funkce, podle níž umíme vzorkovat, vyskytuje v integrantu jako multiplikativní člen (rovnice odrazu, zobrazovací rovnice).
- Řídící funkce
 - Lepší pokud se funkce, kterou umíme analyticky integrovat, vyskytuje v integrantu jako aditivní člen.
- Proto v se v syntéze obrazu téměř vždy používá importance sampling.

Lepší rozmístění vzorků

- Při výběru množiny nezávislých vzorků se stejnou hustotou pravděpodobnosti dochází ke shlukování
 - velký rozptyl odhadu
- Lepší rozmístění vzorků = integrační oblast je pravidelněji pokryta
 - snížení rozptylu
- Metody
 - Vzorkování po částech (stratifikace, **stratified sampling**)
 - quasi-Monte Carlo (QMC)

Vzorkování po částech

- Interval se rozdělí na části, které se odhadují samostatně



Vzorkování po částech

Rozdělení intervalu Ω na N částí Ω_i :

$$I = \int_{\Omega} f(x) dx = \sum_{i=1}^N \int_{\Omega_i} f(x) dx = \sum_{i=1}^N I_i$$

Estimátor:

$$\hat{I}_{\text{strat}} = \frac{1}{N} \sum_{i=1}^N f(X_i), \quad X_i \in \Omega_i$$

Vzorkování po částech

- Potlačuje shlukování vzorků
- Redukuje rozptyl odhadu
 - Rozptyl menší nebo roven rozptylu sekundárního estimátoru
- Velmi účinné pro nízkou dimenzi integrantu

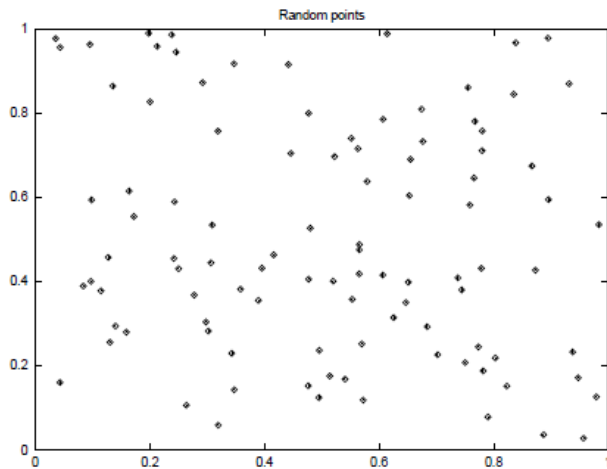
Rozklad intervalu na části

- **uniformní** rozklad intervalu **(0,1)**
 - přirozená metoda pro zcela neznámou funkci **f**
- známe-li alespoň přibližně **průběh funkce f** , snažíme se o takový rozklad, aby byl rozptyl funkce na subintervalech co nejmenší
- rozklad **d -rozměrného intervalu** vede na **N^d** výpočtů
 - úspornější metodou je vzorkování “ **N věží**”

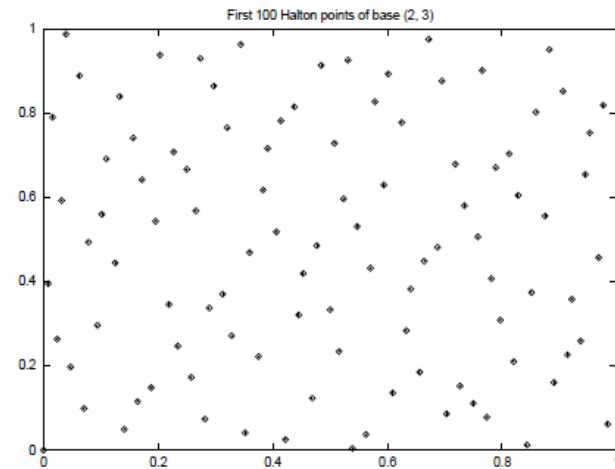
Metody Quasi Monte Carlo (QMC)

- Použití striktně deterministických sekvencí místo náhodných čísel
- Vše funguje jako v MC, důkazy se ale nemohou opírat o statistiku (nic není náhodné)
- Použité sekvence čísel s nízkou dikrepancí (**low-discrepancy sequences**)

Diskrepanz



High Discrepancy
(clusters of points)



Low Discrepancy
(more uniform)

Defining discrepancy

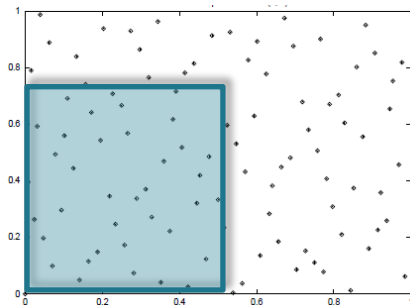
- s -dimensional “brick” function:

$$\mathcal{L}(\mathbf{z}) = \begin{cases} 1 & \text{if } 0 \leq \mathbf{z}|_1 \leq v_1, 0 \leq \mathbf{z}|_2 \leq v_2, \dots, 0 \leq \mathbf{z}|_s \leq v_s \\ 0 & \text{otherwise.} \end{cases}$$

- True volume of the “brick” function:

$$V(A) = \prod_{j=1}^s v_j$$

- MC estimate of the volume of the “brick”:



$$\frac{1}{N} \sum_{i=1}^N f(\mathbf{z}_i) = \frac{m(A)}{N}$$

total number of sample points

number of sample points that actually fell inside the “brick”

Discrepancy

- Discrepancy (of a point sequence) is the maximum possible error of the MC quadrature of the “brick” function over all possible brick shapes:

$$\mathcal{D}^*(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N) = \sup_A \left| \frac{m(A)}{N} - V(A) \right|$$

- serves as a measure of the uniformity of a point set
- must converge to zero as $N \rightarrow \infty$
- the lower the better (cf. **Koksma-Hlawka Inequality**)

Koksma-Hlawka inequality

- Koksma-Hlawka inequality

„variation“ of f

$$\left| \int_{\mathbf{z} \in [0,1]^s} f(\mathbf{z}) \, d\mathbf{z} - \frac{1}{N} \sum_{i=1}^N f(\mathbf{z}_i) \right| \leq \mathcal{V}_{\text{HK}} \cdot \mathcal{D}^*(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$$

- ❑ the KH inequality only applies to f with finite variation
- ❑ QMC can still be applied even if the variation of f is infinite

Van der Corput Sequence

- b ... **base**, must be relative prime (2,3,5,7,...)
- radical inverse

$$\begin{aligned} \Phi_b : \mathbb{N}_0 &\rightarrow \mathbb{Q} \cap [0, 1) \\ i = \sum_{j=0}^{\infty} a_j(i) b^j &\mapsto \Phi_b(i) := \sum_{j=0}^{\infty} a_j(i) b^{-j-1} \end{aligned}$$

Van der Corput Sequence (base 2)

i	binary form of i	radical inverse	H_i
1	1	0.1	0.5
2	10	0.01	0.25
3	11	0.11	0.75
4	100	0.001	0.125
5	101	0.101	0.625
6	110	0.011	0.375
7	111	0.111	0.875

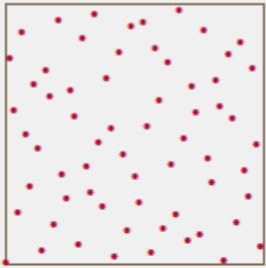
- point placed in the middle of the interval
- then the interval is divided in half
- has low-discrepancy

Van der Corput Sequence (base b)

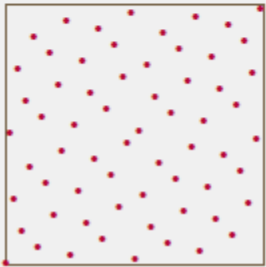
```
double RadicalInverse(const int Base, int i)
{
    double Digit, Radical, Inverse;
    Digit = Radical = 1.0 / (double) Base;
    Inverse = 0.0;
    while(i)
    {
        Inverse += Digit * (double) (i % Base);
        Digit *= Radical;
        i /= Base;
    }
    return Inverse;
}
```

Radical inversion based points in higher dimension

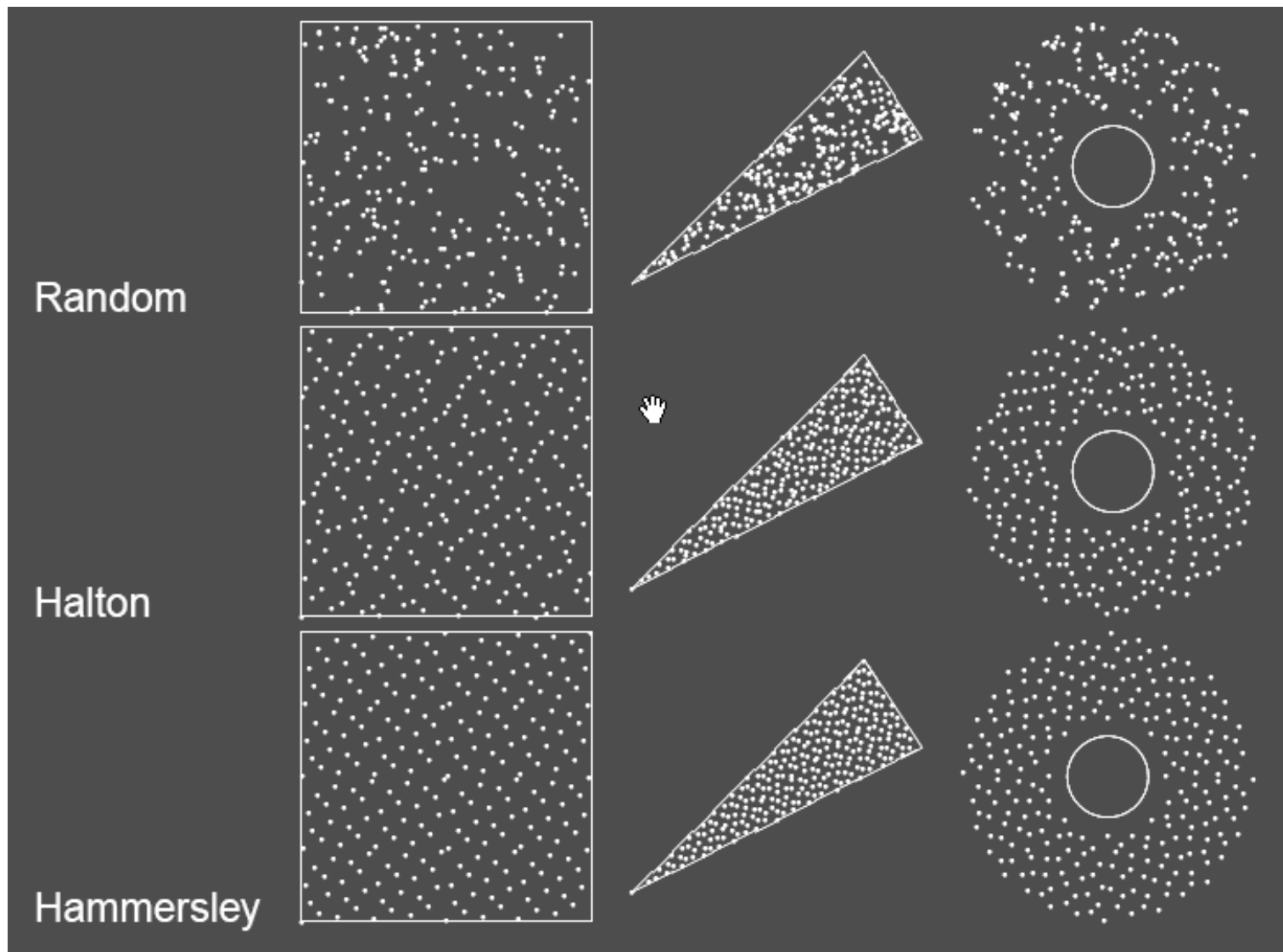
Halton sequence $x_i := (\Phi_{b_1}(i), \dots, \Phi_{b_s}(i))$ where b_i is the i -th prime number



Hammersley point set $x_i := \left(\frac{i}{n}, \Phi_{b_1}(i), \dots, \Phi_{b_{s-1}}(i) \right)$



Transformace náhodných čísel



Ukázka výsledků pro MC a QMC

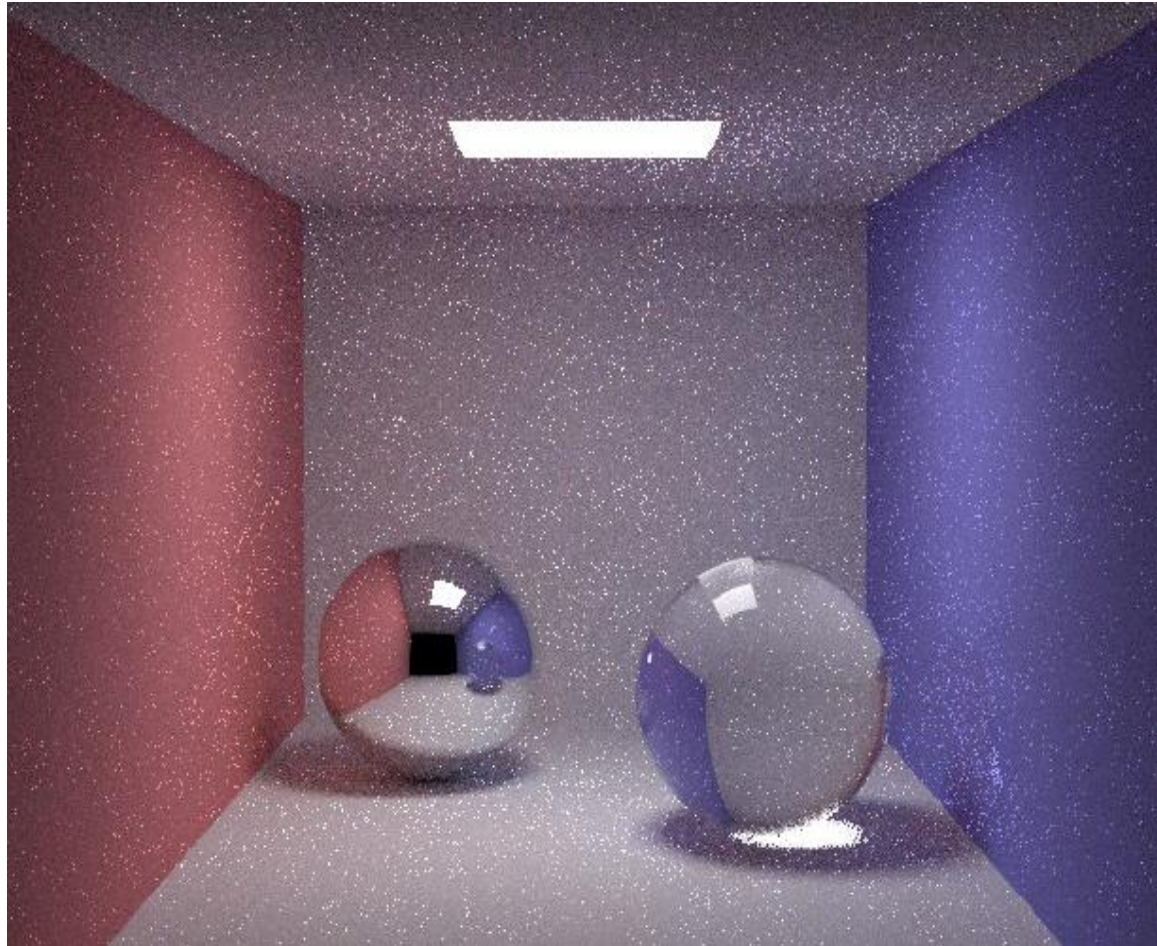
Monte Carlo
(230s)



padded
Hammersley
(202s)



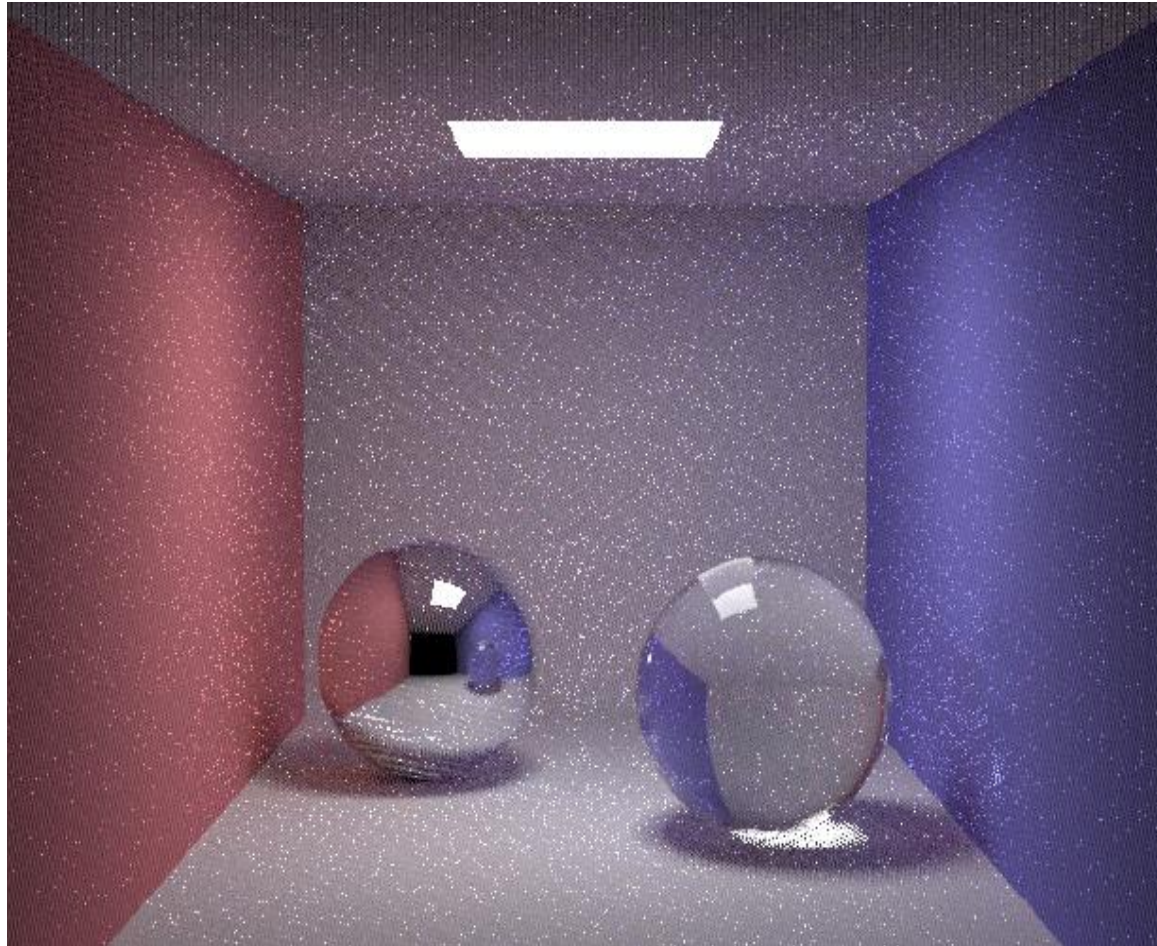
Stratified sampling



10 cest na pixel

Henrik Wann Jensen

Quasi-Monte Carlo



10 cest na pixel

Henrik Wann Jensen

Fixní náhodná sekvence



10 cest na pixel

Henrik Wann Jensen

Metody Quasi Monte Carlo (QMC)

- Nevýhody QMC:
 - V obrázku mohou vzniknout viditelné „vzory“ (místo šumu v MC)