matfyz

Computer Graphics Charles University

Monte Carlo integration

CGG MFF UK Praha

pepca@cgg.mff.cuni.cz

https://cgg.mff.cuni.cz/~pepca/

Monte Carlo integration (quadrature)

Integral to estimate

$$I = \int_0^1 f(x)\, dx$$

Precondition $\quad f(x) \in L^2(0,1)$

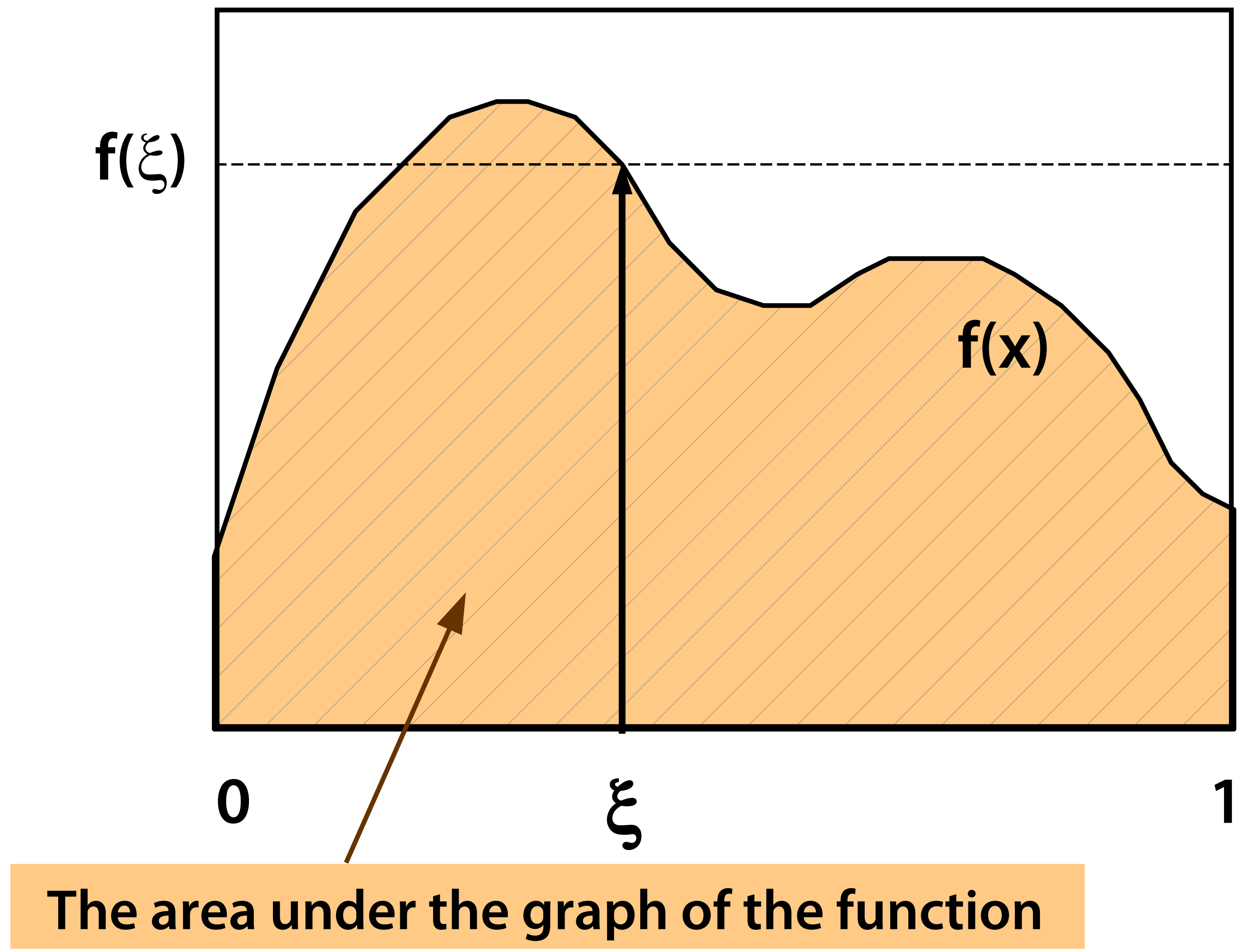
If ξ is a random number from **R(0,1)** distribution, than **f(ξ)** is a **primary estimate** of the integral

$$\langle I \rangle_{prim} = f(\xi)$$

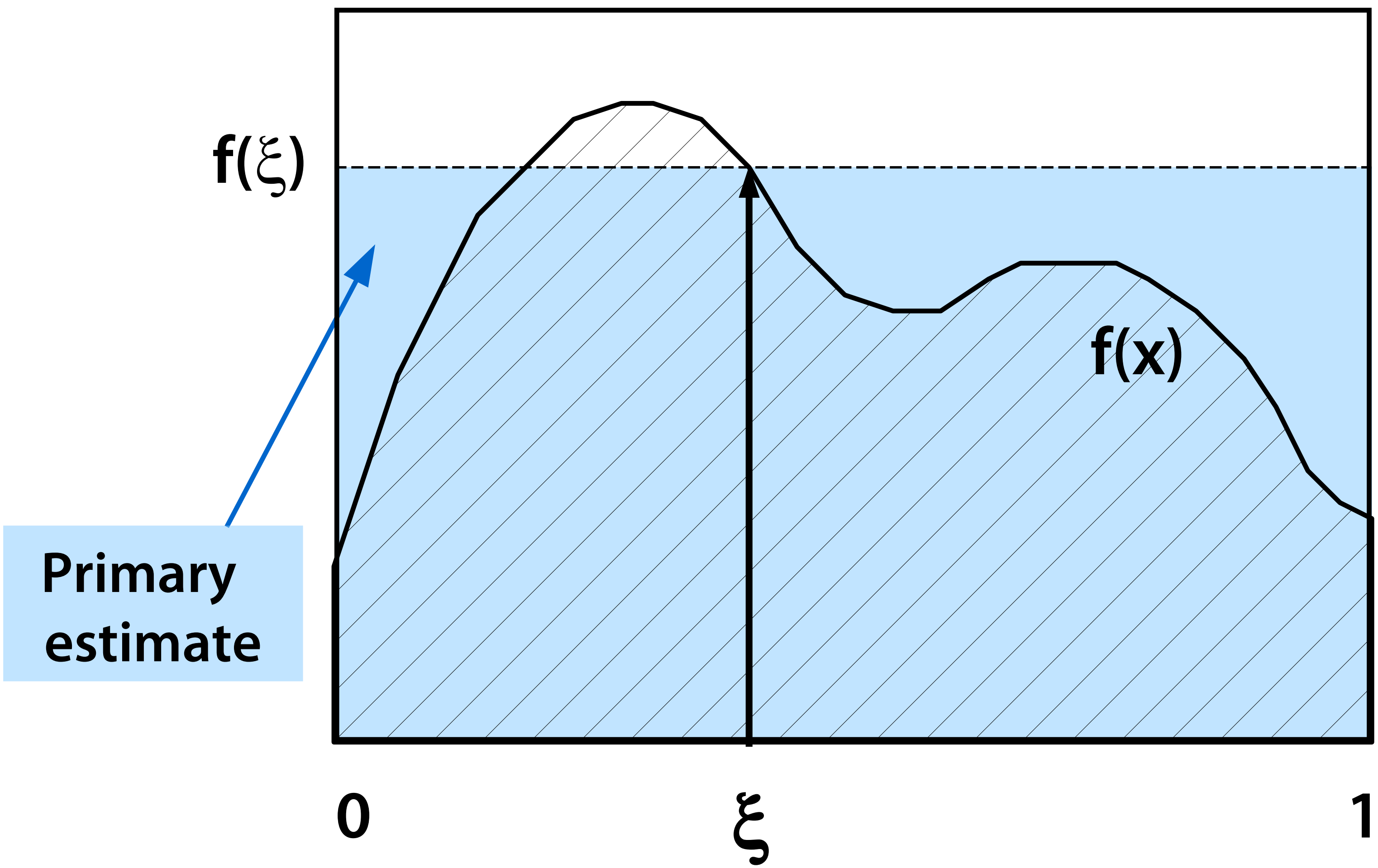
Estimate is **unbiased**, because

$$E\left(\langle I \rangle_{prim}\right) = \int_0^1 f(x)\, dx = I$$

Primary estimate



Estimating the area



Variance of the primary estimate

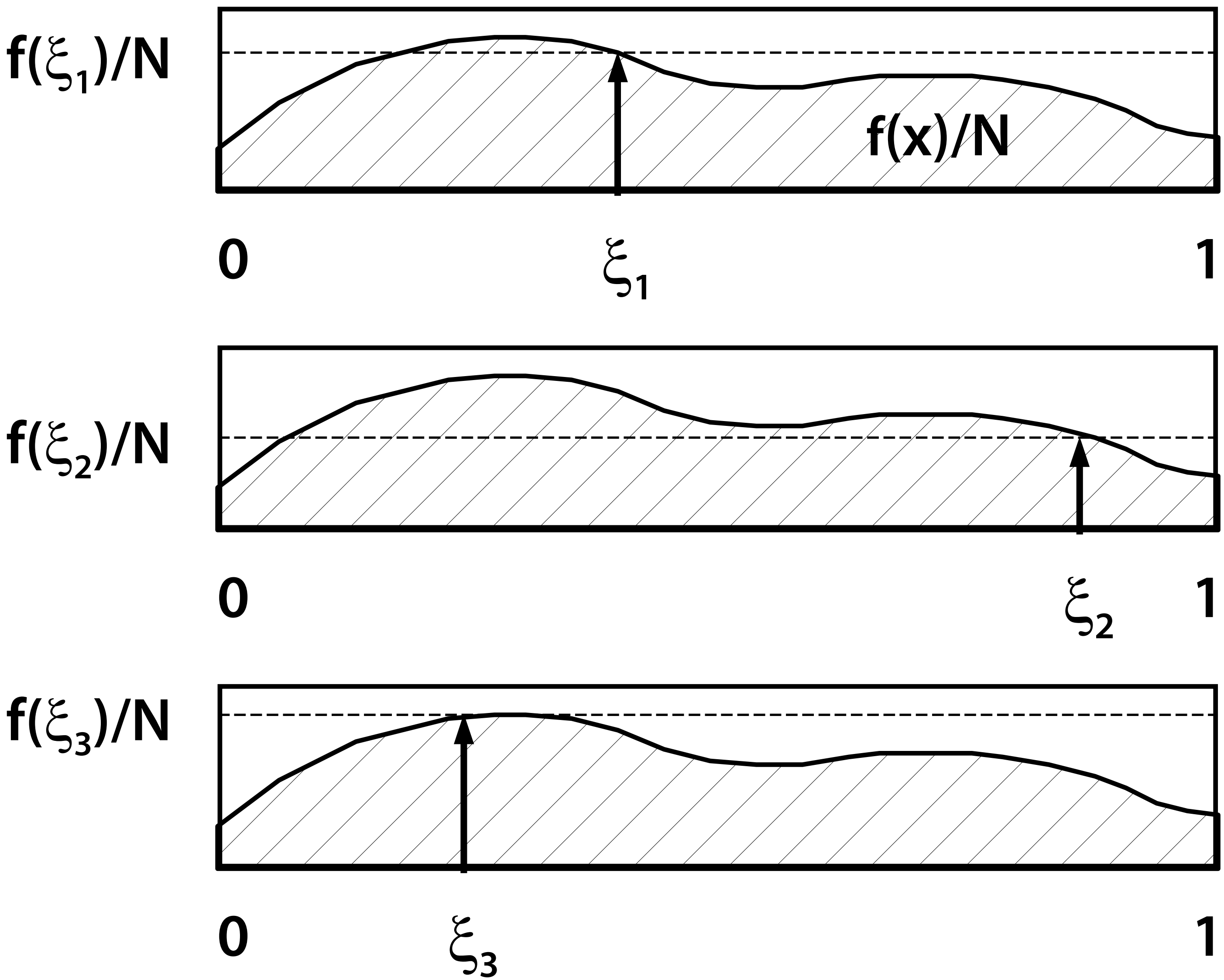
Quality measure of an estimate is the **variance**
(or standard deviation in statistics)

$$V\left(\langle I\rangle_{prim}\right) = \sigma^2_{prim} = \int_0^1 \left| f(x) - I \right|^2 dx = \int_0^1 f(x)^2 dx - I^2$$

(for unbiased estimate)

If we have a **single sample only**, the variance is too high!

Secondary estimate



Secodary estimate

The integral is divided into **N addends**

$$I = \int_0^1 f(x)\,dx = \sum_{i=1}^{N} \int_0^1 \frac{f(x)}{N}\,dx = \sum_{i=1}^{N} I_i$$

Secondary estimate of the integral

$$\langle I \rangle_{sec} = \sum_{i=1}^{N} \langle I_i \rangle_{prim} = \frac{1}{N} \sum_{i=1}^{N} f(\xi_i)$$

Secondary estimate is **unbiased** as well

Variance of the secondary estimate

$$\sigma^2_{sec} = \int_0^1 \cdots \int_0^1 \left[\frac{1}{N} \sum_{i=1}^{N} f(x_i) \right]^2 dx_1 \ldots dx_N - I^2 =$$

$$= \frac{1}{N} \int_0^1 f^2(x)\, dx - \frac{1}{N} I^2 =$$

$$= \frac{\sigma^2_{prim}}{N}$$

… variance is √**N-times smaller!**
(convergence: **1/√N**)

Stratified sampling

If we are sampling the whole set using constant distribution, **clustering is inevitable**

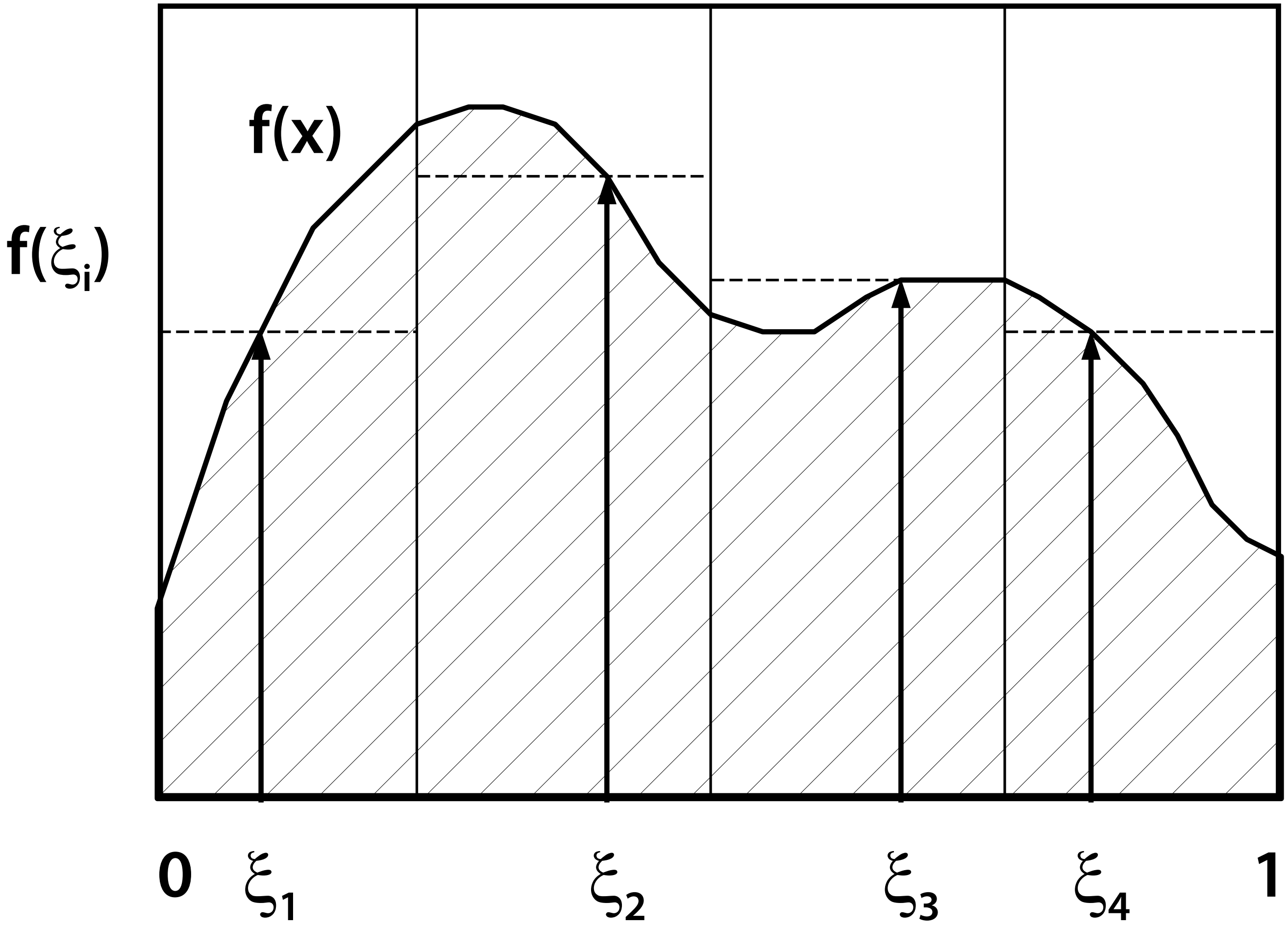
- too high variance

Stratified sampling

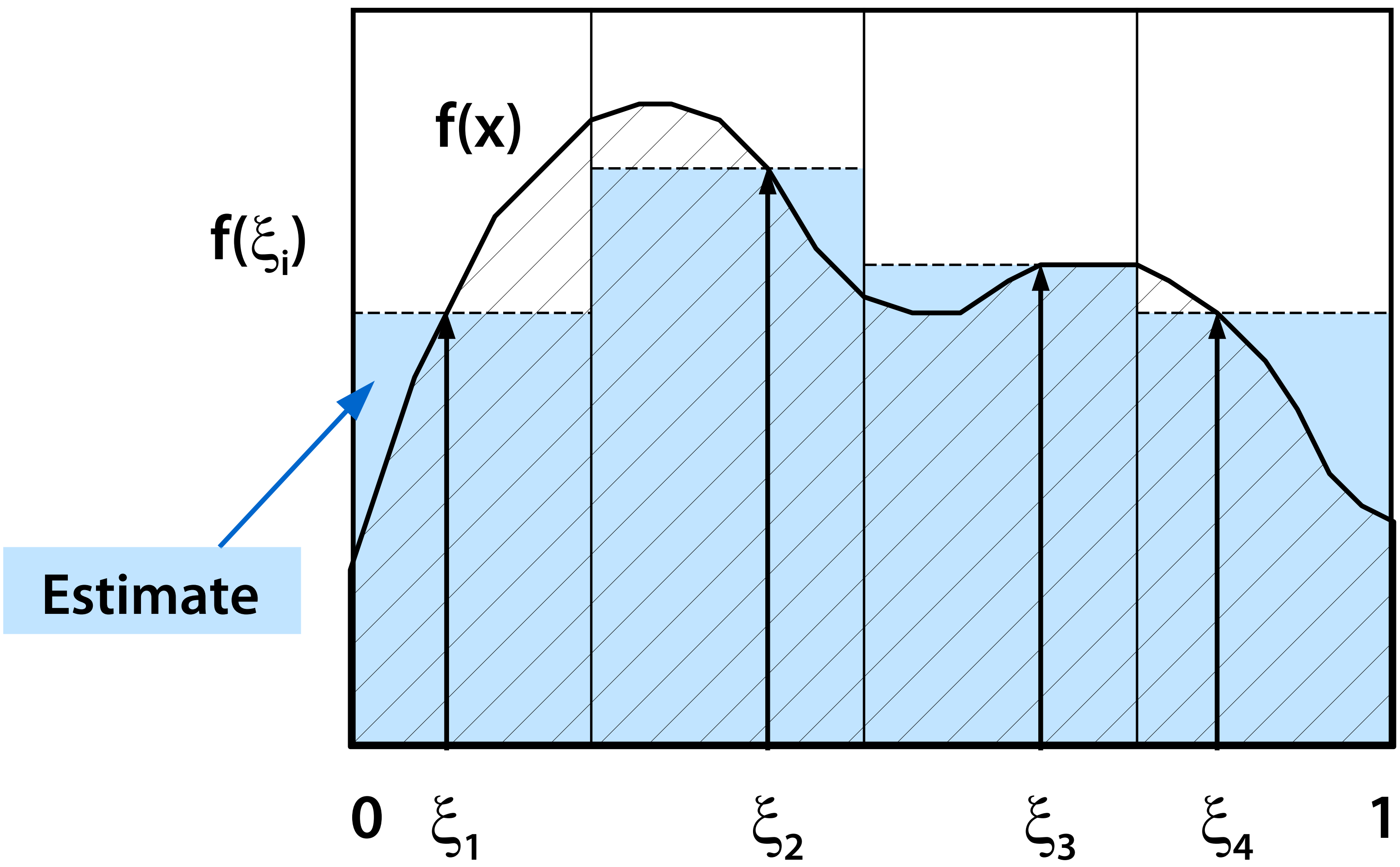
- reduces clustering
- reduces variance
- very popular in sampling discrete sets as well (statistics in sociology, etc.)

The domain is divided into intervals which are then sampled **independently**

Stratified sampling



Stratified sampling – areas



Stratified sampling

Domain **(0,1)** is divided **into N parts A_i**

$$I = \int_0^1 f(x)\,dx = \sum_{i=1}^{N} \int_{A_i} f(x)\,dx = \sum_{i=1}^{N} I_i$$

Estimate

$$\langle I \rangle_{strat} = \sum_{i=1}^{N} \langle I_i \rangle_{prim} = \frac{1}{N} \sum_{i=1}^{N} f(\xi_i), \qquad f(\xi_i) \in A_i$$

Variance of the stratified sampling

$$\sigma^2_{strat} = \sum_{i=1}^{N} \left[\int_{A_i} \left[\frac{f(x_i)}{N} \right]^2 N \, dx_i - I_i^2 \right] =$$

$$= \frac{1}{N} \int_0^1 f^2(x) \, dx - \sum_{i=1}^{N} I_i^2 \leq \sigma^2_{sec}$$

Meaning – variance **cannot be higher** than variance of the **secondary estimate**!

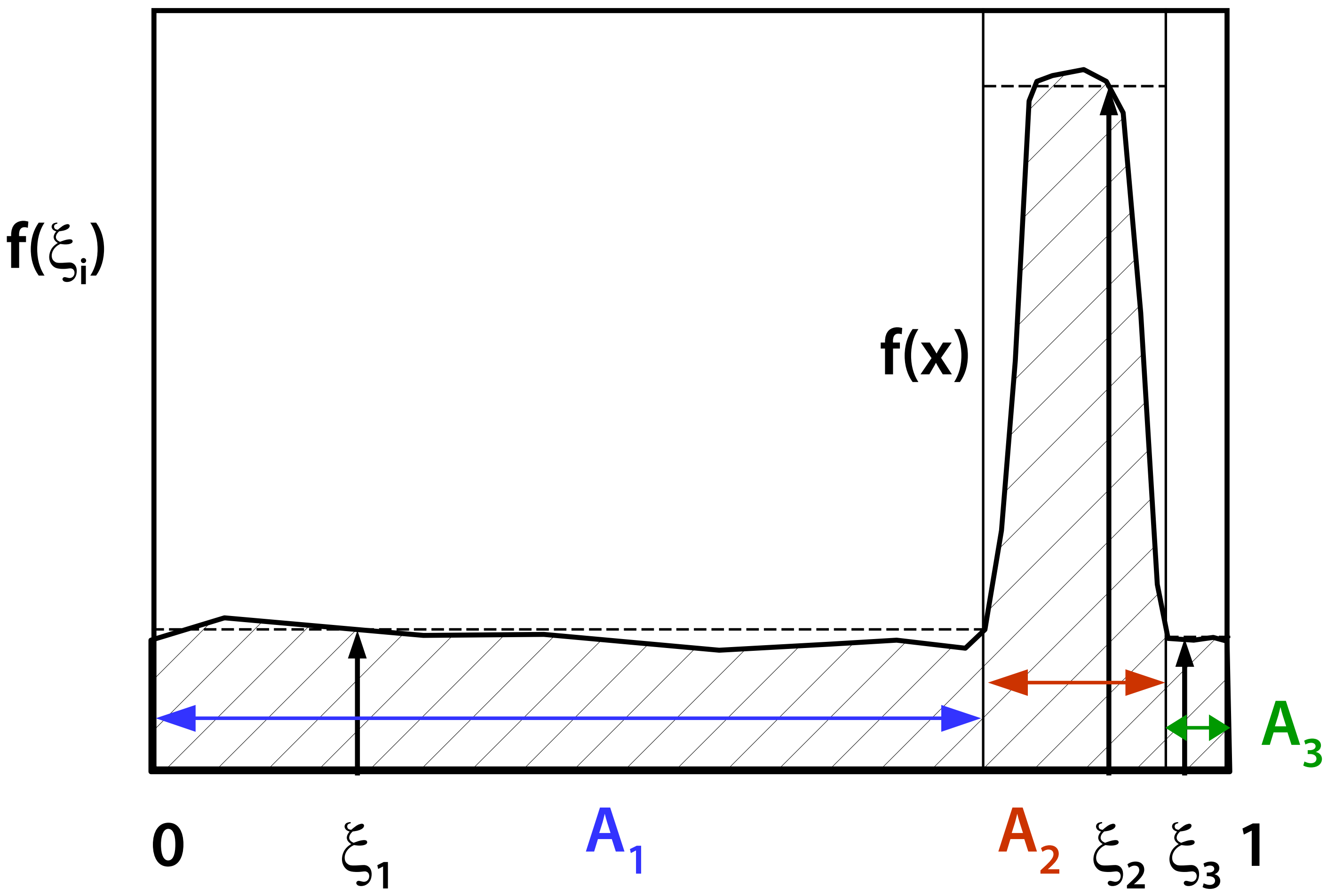
Division of the domain

Uniform division of the interval **(0,1)**
- natural approach for unknown function **f**

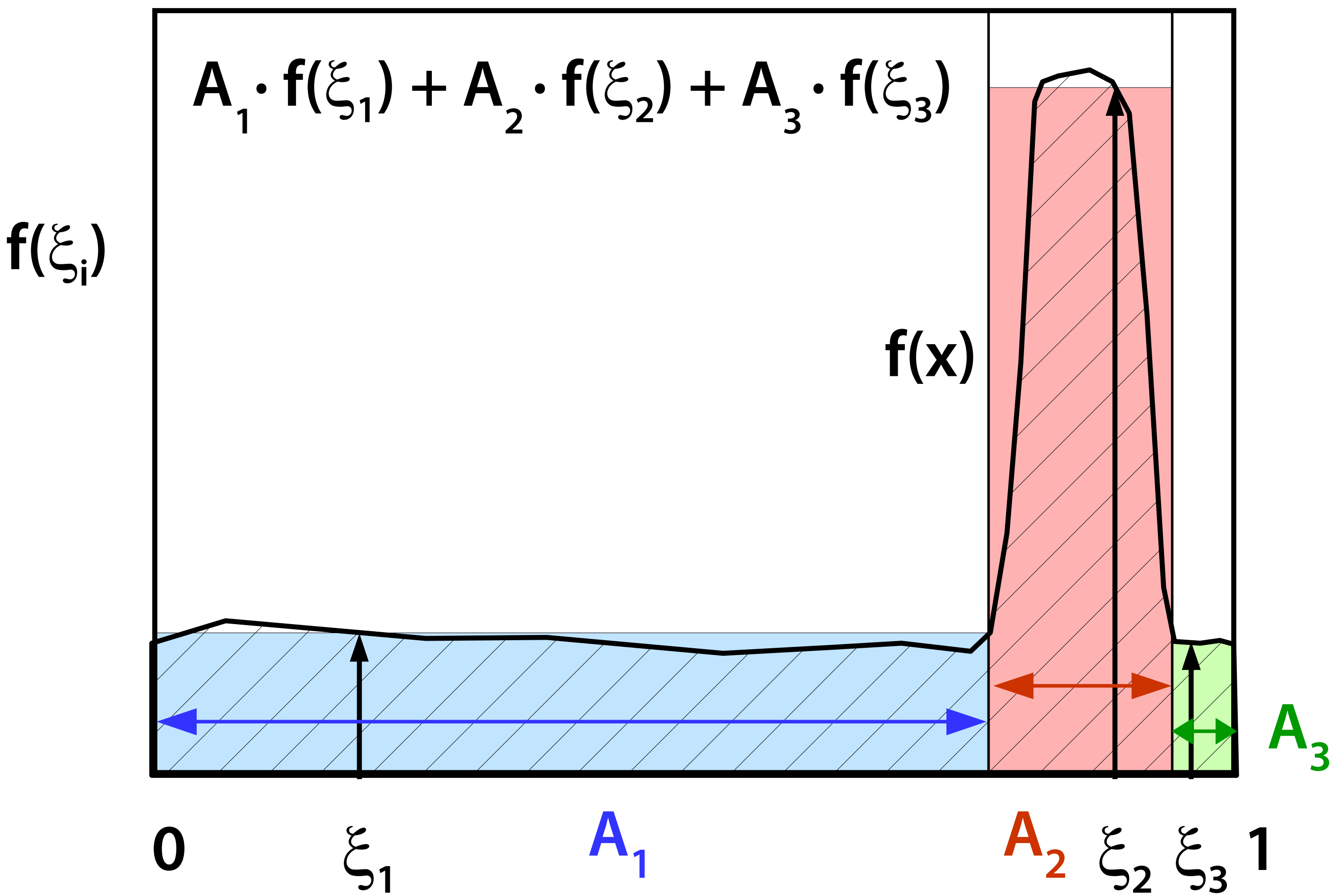
If we know more about **f waveform**, our goal is to use a division induced by a shape of the function (to minimize partial variances)

Division of **d-dimensional interval** leads to $\mathbf{N^d}$ samples
- the **“N rooks”** sampling is smarter

High variance function



Estimate for smarter domain division



Importance sampling

Sampling is more important in parts with higher value of the function **f**

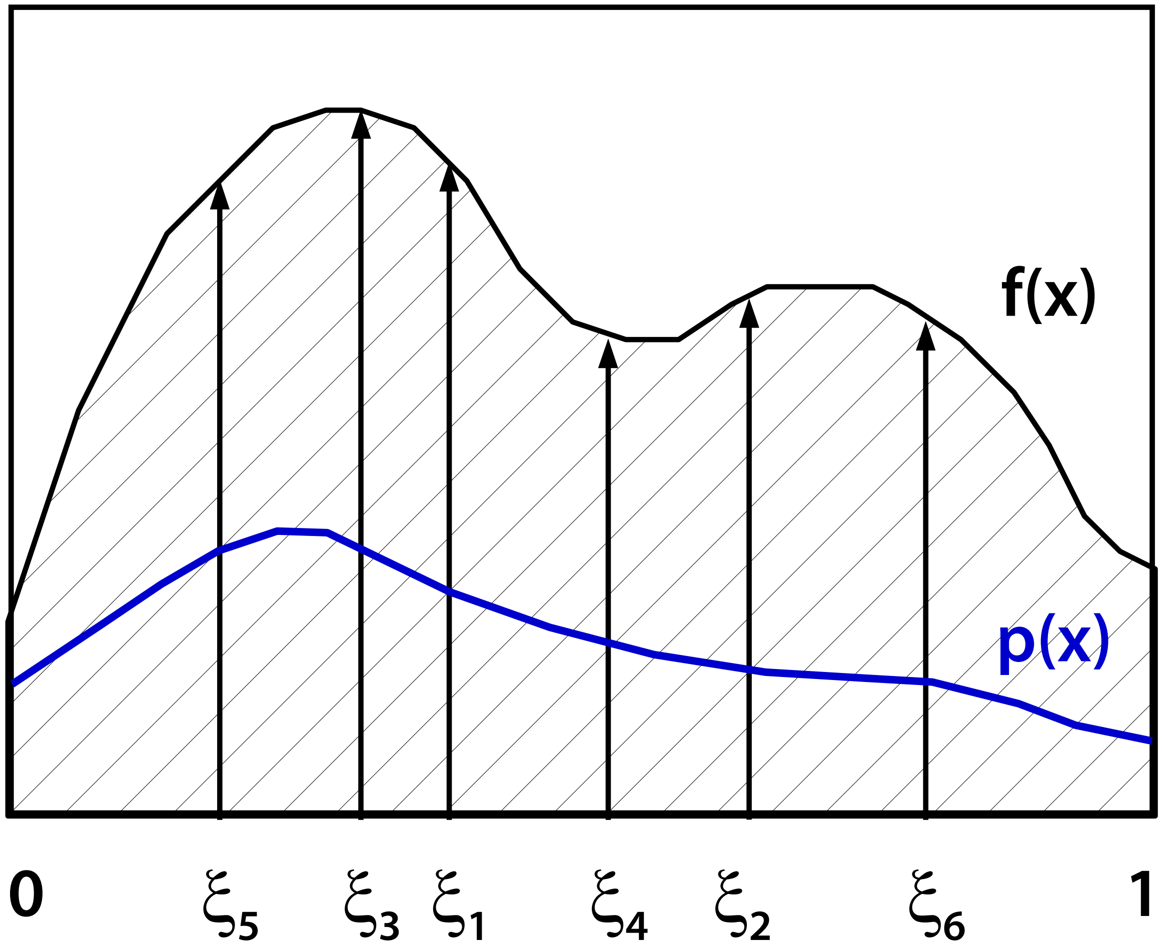
- samples from that regions have higher impact (relative contribution to the result)

The **“importance sampling”** prioritizes such samples

- formally: sampling is controlled by custom **probability distribution function** (density) **p(x)**

Lower variance but still an unbiased estimate

Importance sampling



Importance sampling

Estimated integral is transformed

$$I = \int_0^1 f(x)\,dx = \int_0^1 \frac{f(x)}{p(x)}\,p(x)\,dx$$

If PDF of the random variable ξ is **p(x)**, we use an estimate

$$\langle I \rangle_{imp} = \frac{f(\xi)}{p(\xi)}$$

(unbiased)

Variance of the importance sampling

$$\sigma^2_{imp} = \int_0^1 \left[\frac{f(x)}{p(x)}\right]^2 p(x)\,dx - I^2 =$$

$$= \int_0^1 \frac{f^2(x)}{p(x)}\,dx - I^2$$

If the PDF **p(x)** is similar to the integrand **f(x)**, we are estimating integral of function with **lower variance** than **f(x)**

PDF (density) p(x)

PDF is a real function $\mathbf{p(x)} \geq \mathbf{0}$, $\mathbf{p(x)} > \mathbf{0}$ where $\mathbf{f(x)} \neq \mathbf{0}$

Integral $\int \mathbf{p(x)\,dx = 1}$

It is possible to **efficiently generate** samples from the distribution

- the easiest way: cummulative distribution function **P(x)** is known and invertible (we can express $\mathbf{P^{-1}(x)}$)

$$\mathbf{P(x) = \int_0^x p(t)\,dt}$$

Algorithm

Instead of sampling from **p(x),** we compute τ from **the uniform distribution R(0,1)** and transform it

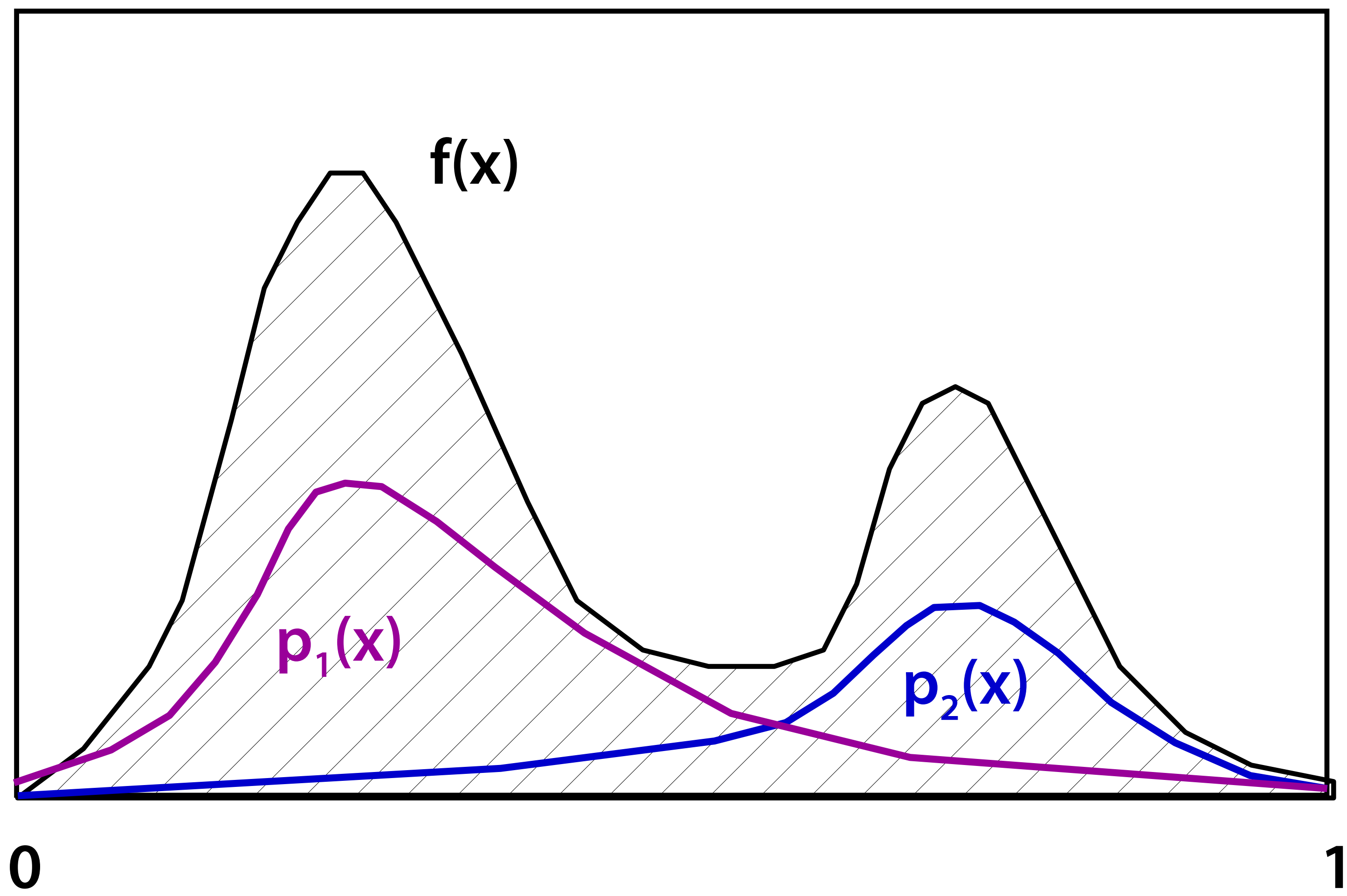
$$\xi = P^{-1}(\tau)$$

Estimate

$$\langle I \rangle_{imp} = \frac{f\left(P^{-1}(\tau)\right)}{p\left(P^{-1}(\tau)\right)}$$

$$I = \int_0^1 f(x)\,dx = \int_0^1 f\left(P^{-1}(t)\right) \frac{dP^{-1}(t)}{dt}\,dt = \int_0^1 \frac{f\left(P^{-1}(t)\right)}{p\left(P^{-1}(t)\right)}\,dt$$

Combined sampling



Combined sampling

Presumption – **n** random variables $\xi_1, .. \xi_n$

with PDFs (densities) $\mathbf{p_1(x), \dots p_n(x)}$

The **combined estimate** of the integral is

$$\langle I \rangle_{comb} = \sum_{i=1}^{n} w_i(\xi_i) \frac{f(\xi_i)}{p_i(\xi_i)}$$

where $\mathbf{w_i(x)}$ are nonnegative **weight functions**

Eliminating bias

$$E\left(\langle I\rangle_{comb}\right) = \sum_{i=1}^{n} \int_0^1 \left[w_i(x_i) \frac{f(x_i)}{p_i(x_i)} \right] p_i(x_i)\, dx_i =$$

$$= \int_0^1 \left[\sum_{i=1}^{n} w_i(x) \right] f(x)\, dx \equiv \int_0^1 f(x)\, dx$$

The only condition

$$\forall x: \quad \sum_{i=1}^{n} w_i(x) = 1$$

Variance of the combined estimate

$$\sigma^2_{comb} = \sum_{i=1}^{n} \left\{ \begin{array}{l} \displaystyle\int_0^1 \left[w_i(x_i) \frac{f(x_i)}{p_i(x_i)} \right]^2 p_i(x_i)\, dx_i - \\ \displaystyle - \left[\int_0^1 w_i(x_i) \frac{f(x_i)}{p_i(x_i)} p_i(x_i)\, dx_i \right]^2 \end{array} \right\} =$$

$$= \int_0^1 \left[\sum_{i=1}^{n} \frac{w_i^2(x)}{p_i(x)} \right] f(x)\, dx - \sum_{i=1}^{n} \left[\int_0^1 w_i(x)\, f(x)\, dx \right]^2$$

Mean, maximum (cutoff)

$$w_i(x) = \frac{1}{n}$$

$$\langle I \rangle_{average} = \frac{1}{n} \sum_{i=1}^{n} \frac{f(\xi_i)}{p_i(\xi_i)}$$

$$w_i(x) = \begin{cases} 1 & \text{for } p_i(x) = \max_j \{ p_j(x) \} \\ 0 & \text{otherwise} \end{cases}$$

$$\langle I \rangle_{max} = \sum_{i=1}^{n} \left(p_i(\xi_i) = \max_j \{ p_j(\xi_i) \} \right) ? \frac{f(\xi_i)}{p_i(\xi_i)} : 0$$

Balance heuristics

$$w_i(x) = \frac{p_i(x)}{\sum_{j=1}^{n} p_j(x)}$$

$$\langle I \rangle_{bal} = \sum_{i=1}^{n} \frac{f(\xi_i)}{\sum_{j=1}^{n} p_j(\xi_i)}$$

$$\sigma_{bal}^2 = \int_0^1 \frac{f^2(x)}{\sum_{i=1}^{n} p_i(x)} \, dx - \sum_{i=1}^{n} \left[\int_0^1 \frac{p_i(x)}{\sum_{j=1}^{n} p_j(x)} f(x) \, dx \right]^2$$

$$\sigma_{comb}^2 \geq \sigma_{bal}^2 - \left(1 - \tfrac{1}{n}\right) \cdot I^2$$

Power heuristics

Generalization $$w_i(x) = \frac{p_i^\beta(x)}{\sum_{j=1}^{n} p_j^\beta(x)}$$

$$\langle I \rangle_{power} = \sum_{i=1}^{n} \frac{p_i^{\beta-1}(\xi_i)}{\sum_{j=1}^{n} p_j^\beta(\xi_i)} f(\xi_i)$$

$\beta = 1$ … balance, $\beta = \infty$ … maximum

Integrand transformation

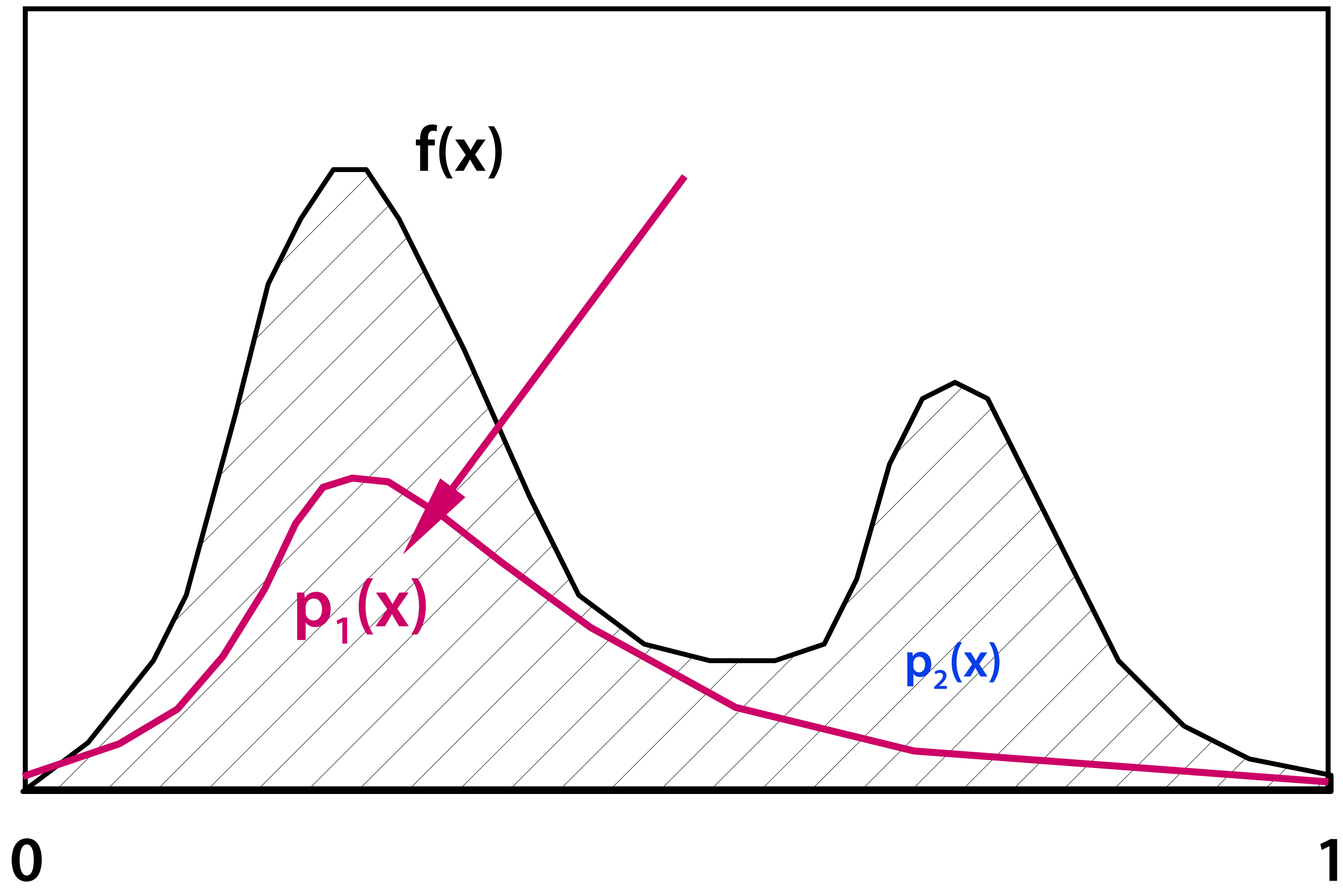
We can interpret the combined estimate as
a transformation of the integrand

$$I = \int_0^1 f(x)\,dx = \sum_{i=1}^{n} \int_0^1 w_i(x) \cdot f(x)\,dx$$

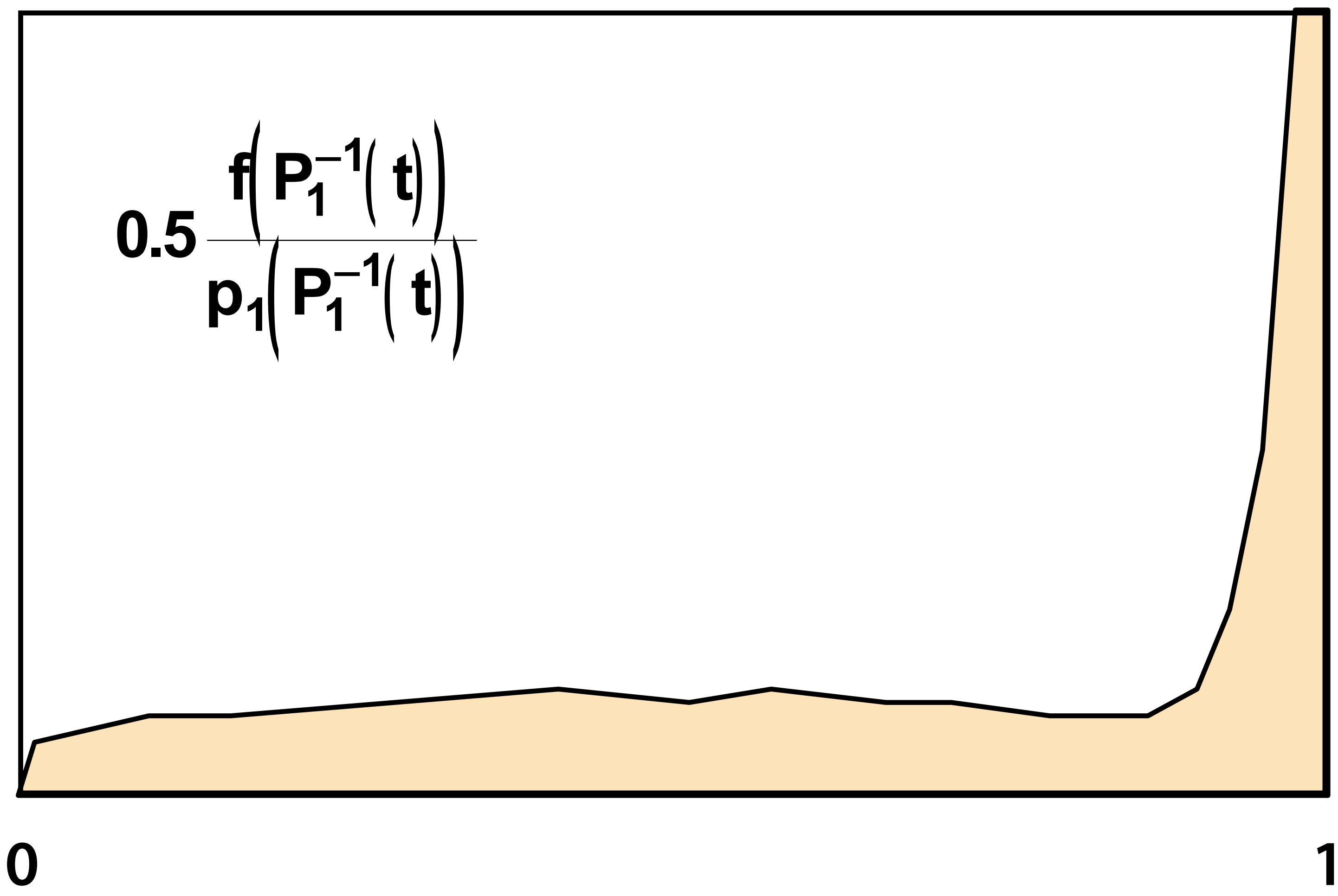
Combination of importance samplings

$$I = \sum_{i=1}^{n} \int_0^1 \frac{w_i\left(P_i^{-1}(t)\right)}{p_i\left(P_i^{-1}(t)\right)} f\left(P_i^{-1}(t)\right)\,dt$$

Example – one term of a combination

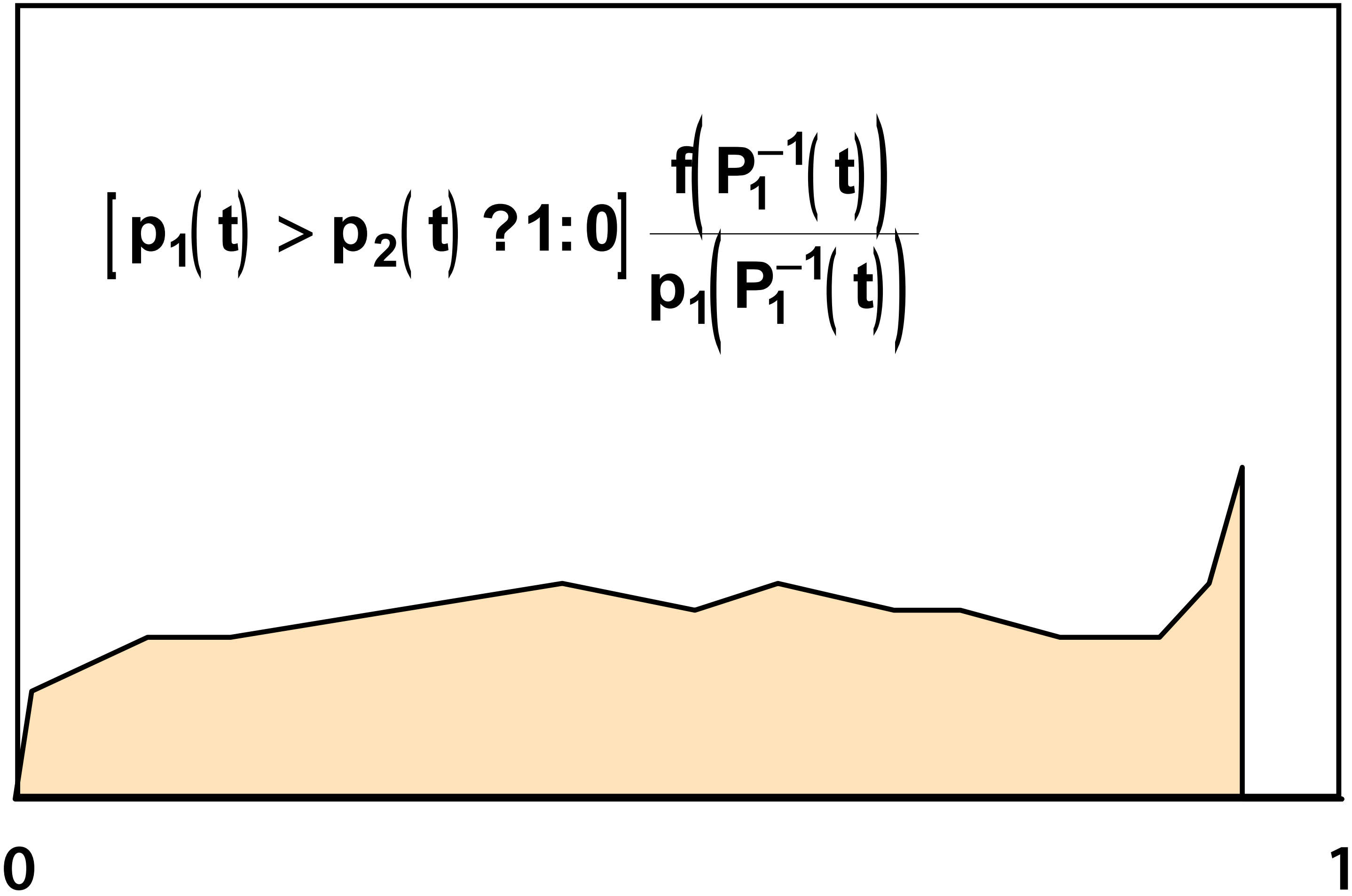


Arithmetic average (mean)



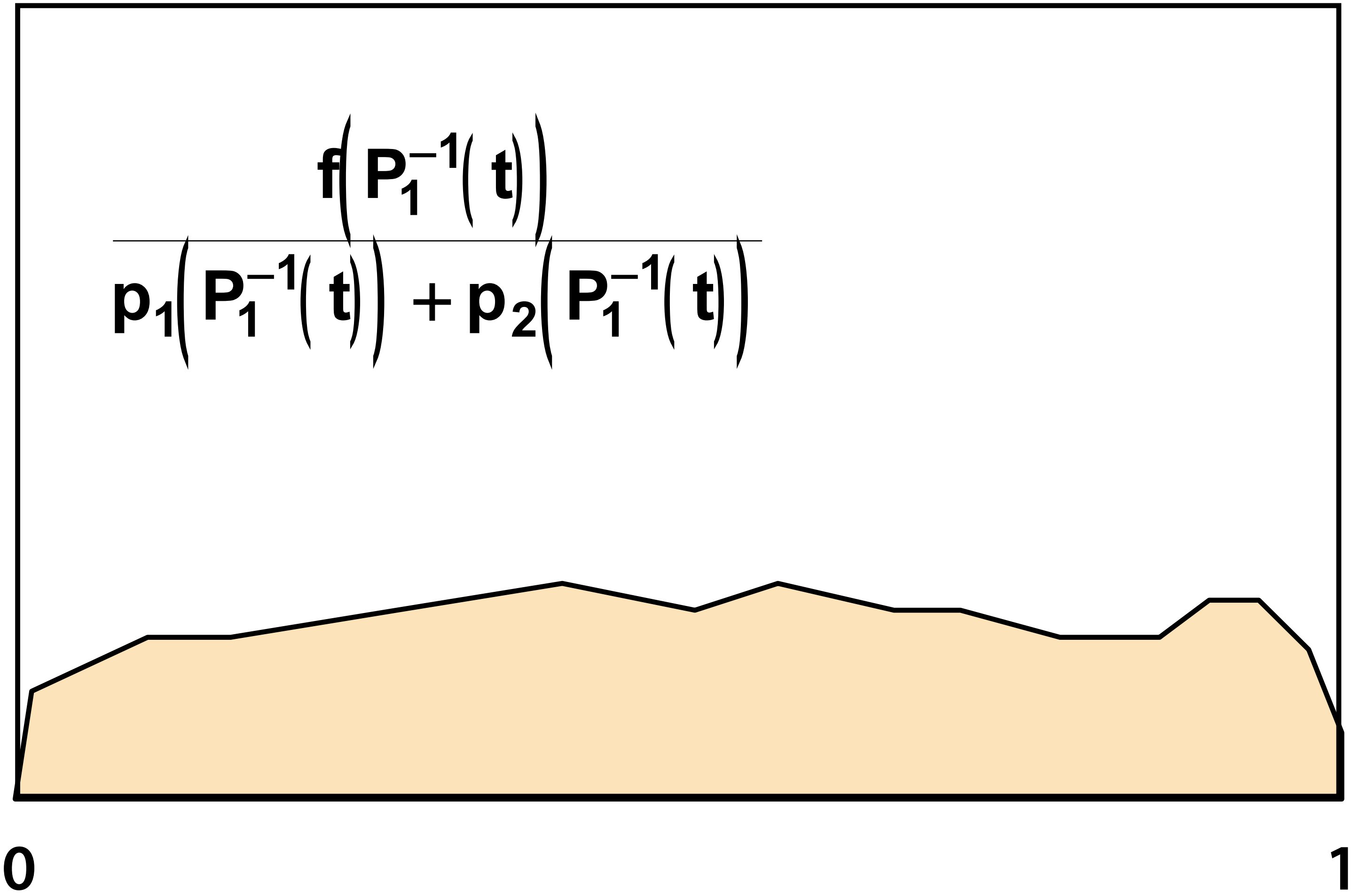
Maximum

$$\left[p_1(t) > p_2(t)\ ?\ 1 : 0 \right] \frac{f\left(P_1^{-1}(t)\right)}{p_1\left(P_1^{-1}(t)\right)}$$



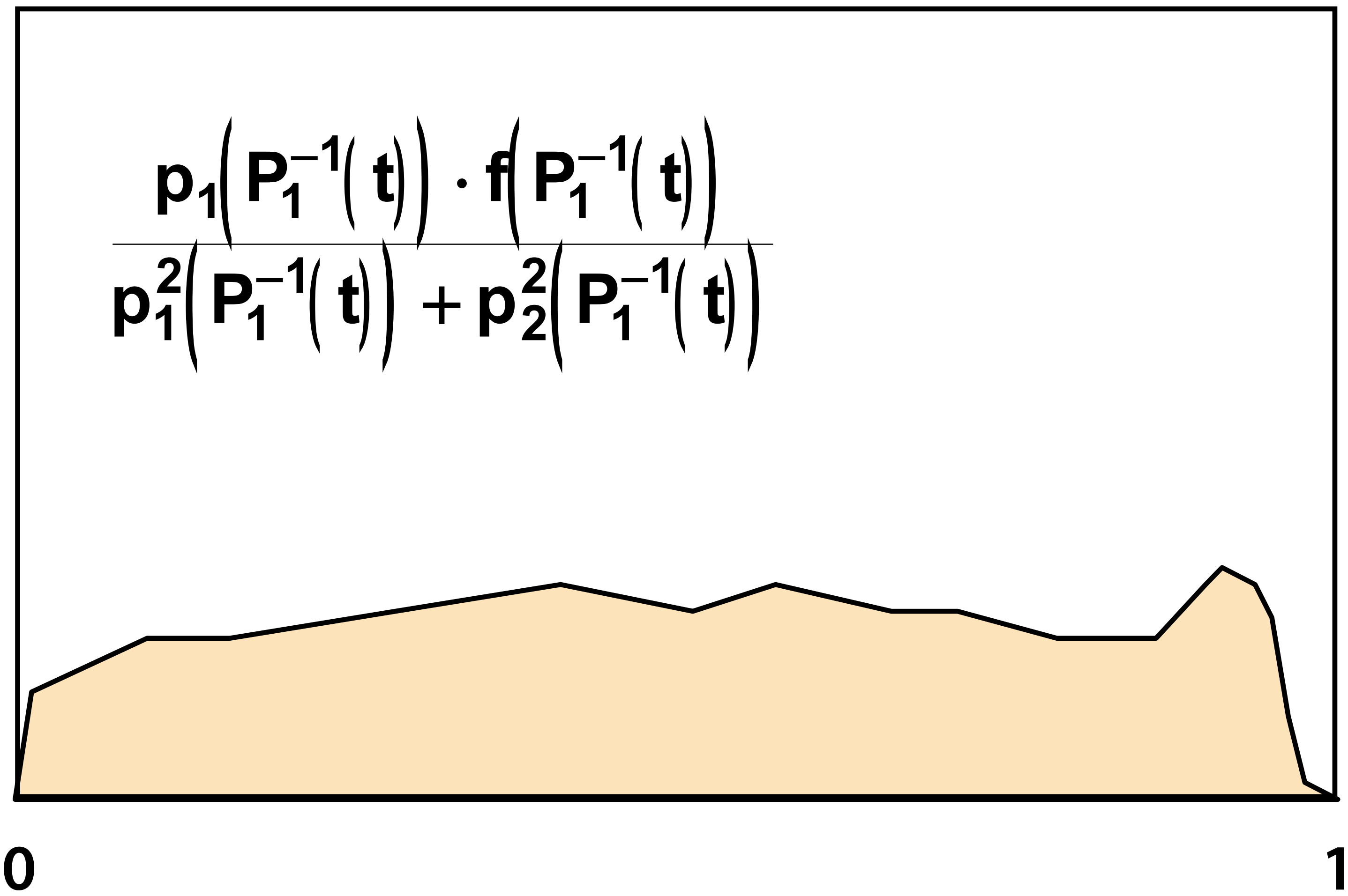
Balance heuristics

$$\frac{f\left(P_1^{-1}(t)\right)}{p_1\left(P_1^{-1}(t)\right)+p_2\left(P_1^{-1}(t)\right)}$$

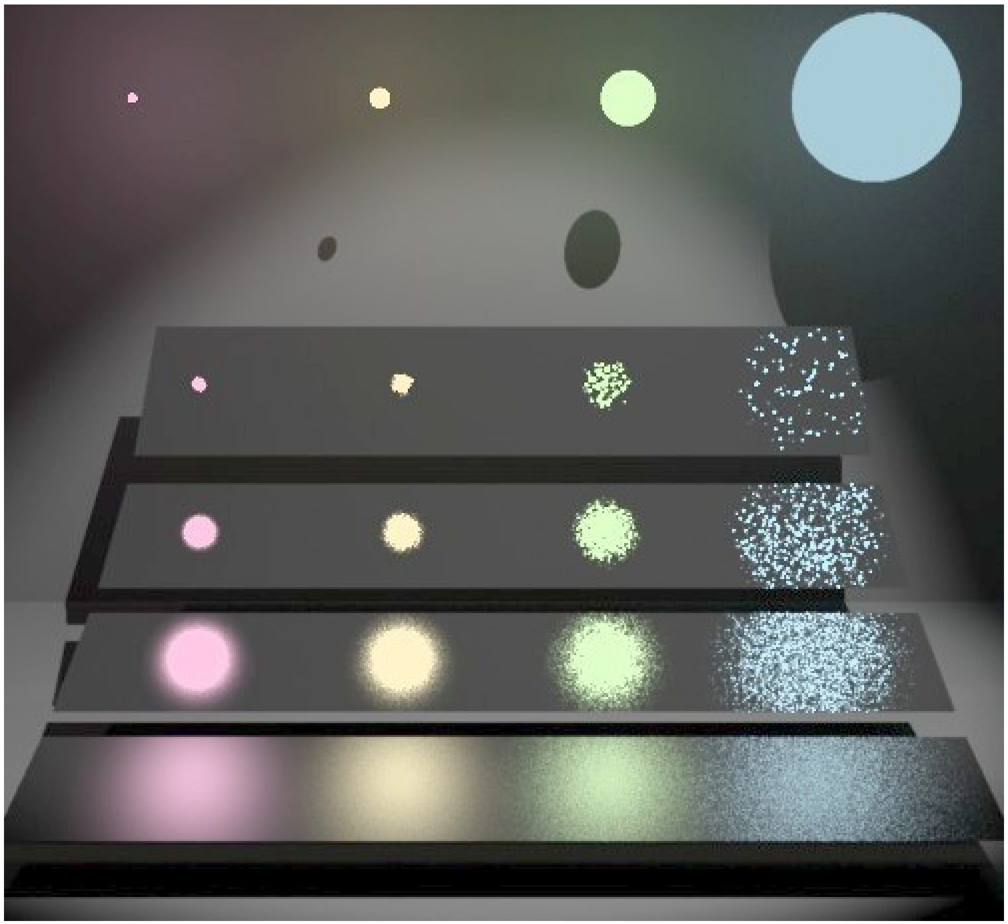


Power heuristics for β=2

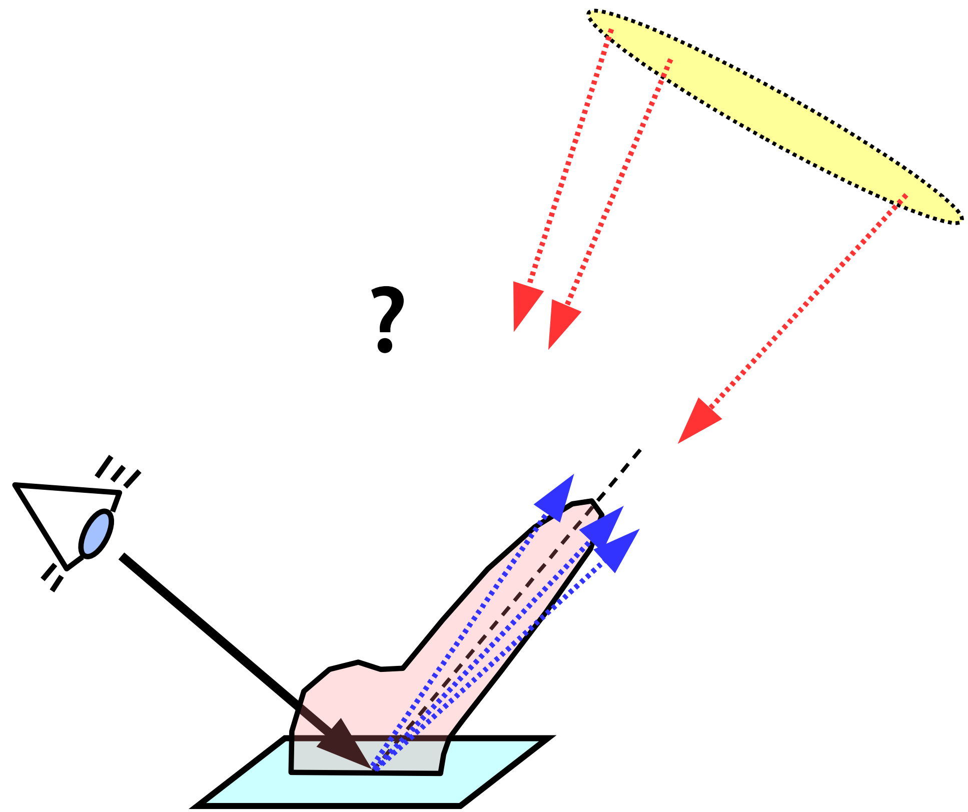
$$\frac{p_1\left(P_1^{-1}(t)\right)\cdot f\left(P_1^{-1}(t)\right)}{p_1^2\left(P_1^{-1}(t)\right)+p_2^2\left(P_1^{-1}(t)\right)}$$



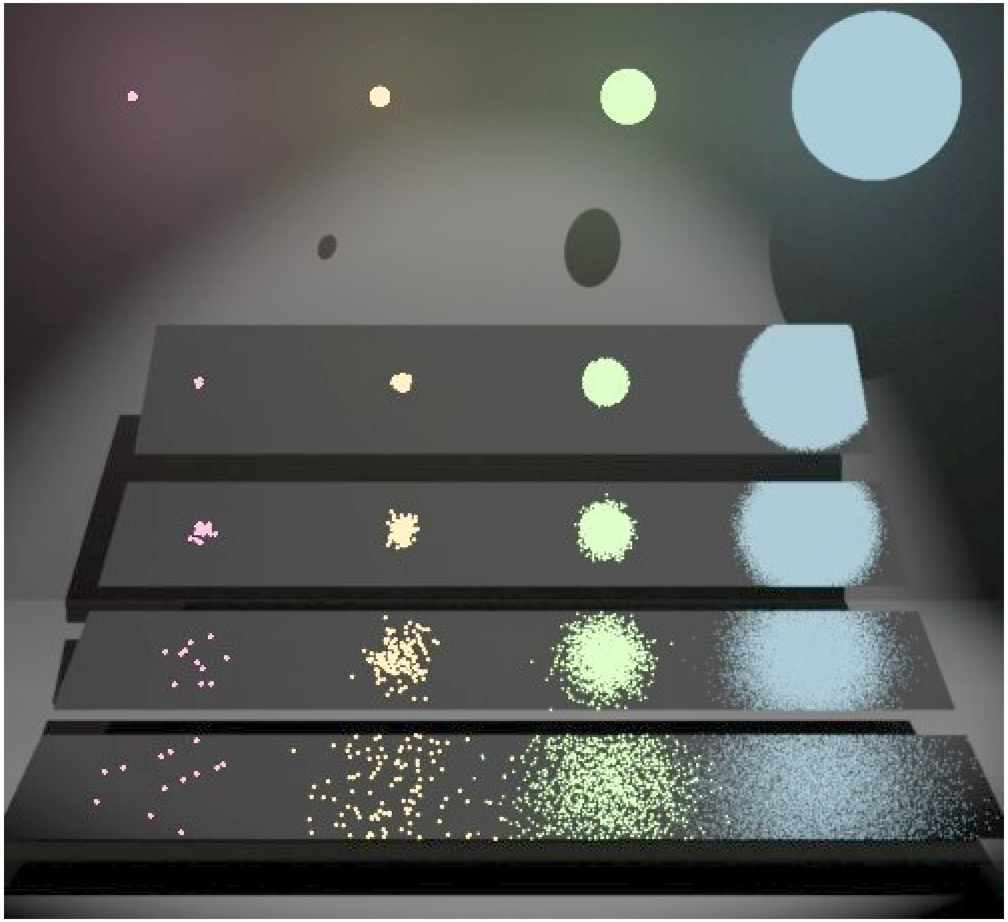
Rendering example



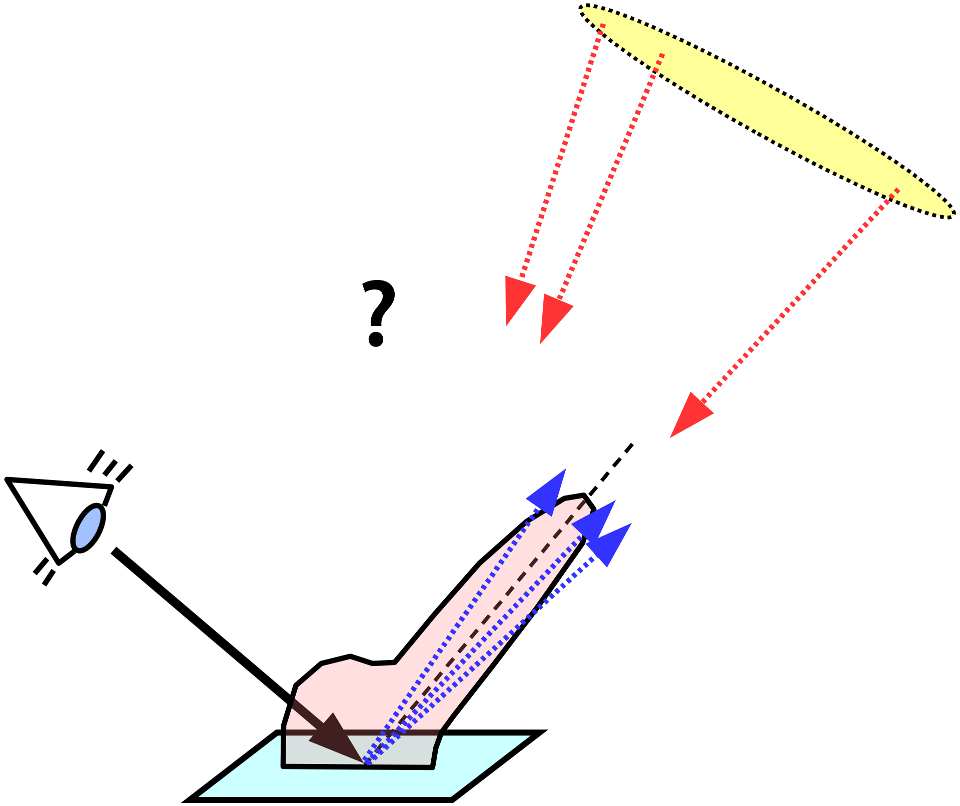
© 1995 Eric Veach, Leonidas J. Guibas



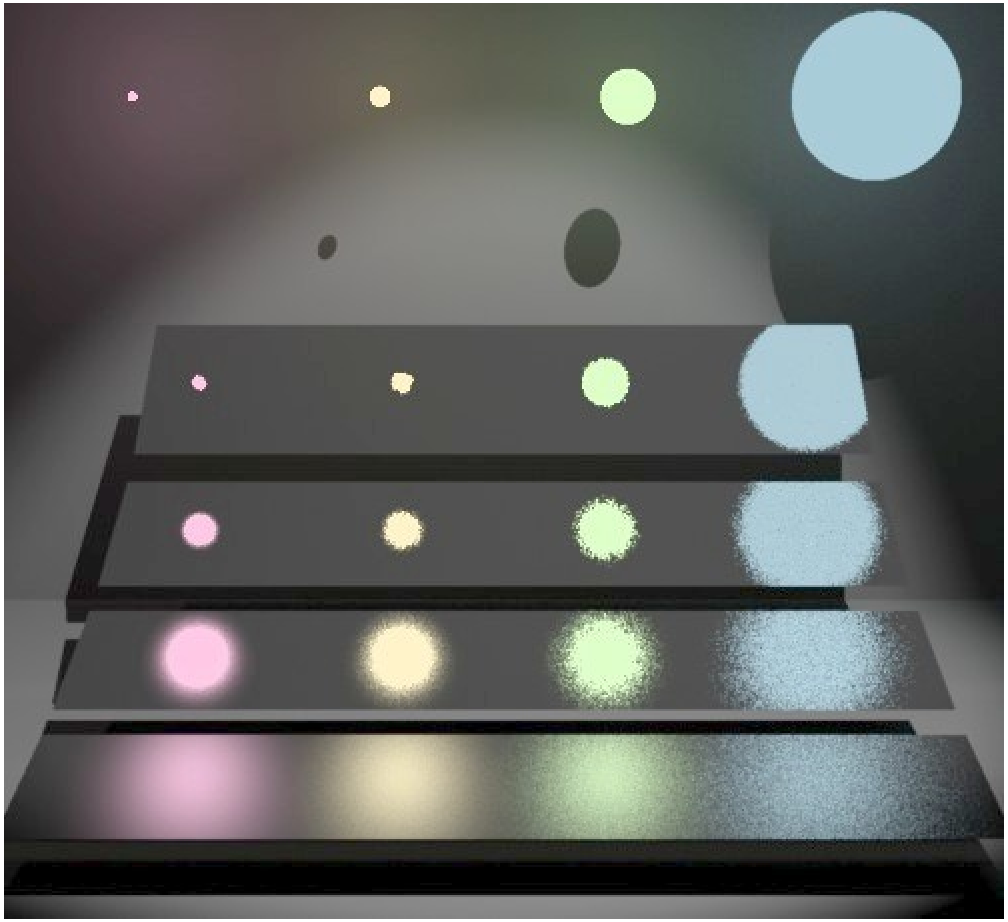
Rendering example



© 1995 Eric Veach, Leonidas J. Guibas



Combined sampling

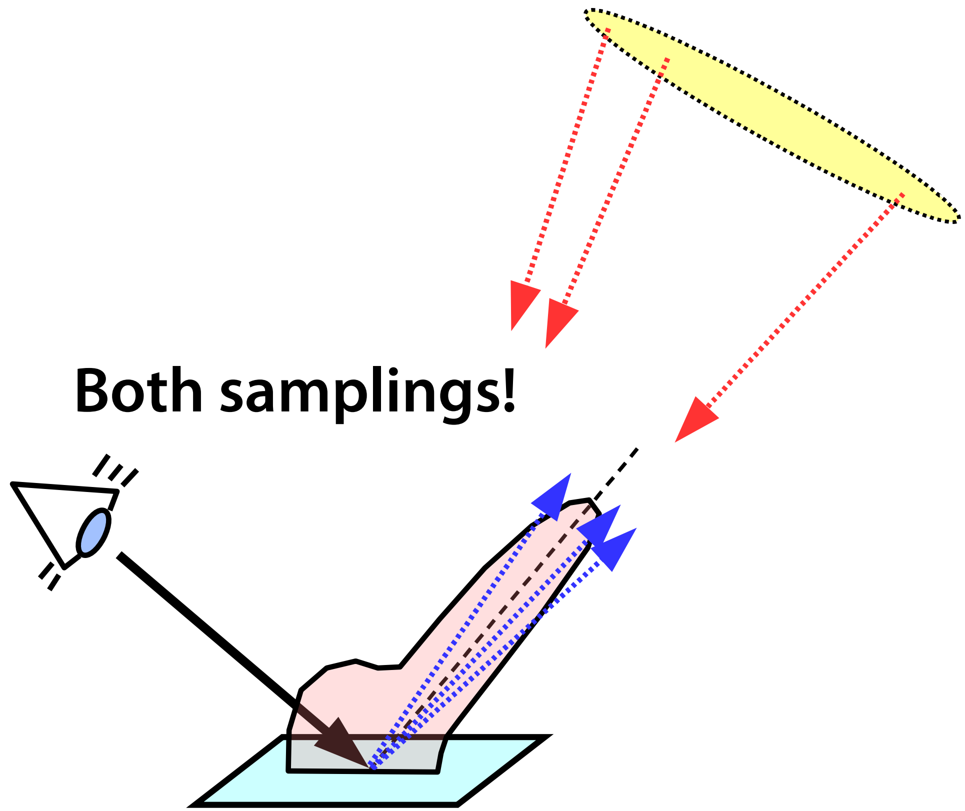


© 1995 Eric Veach, Leonidas J. Guibas

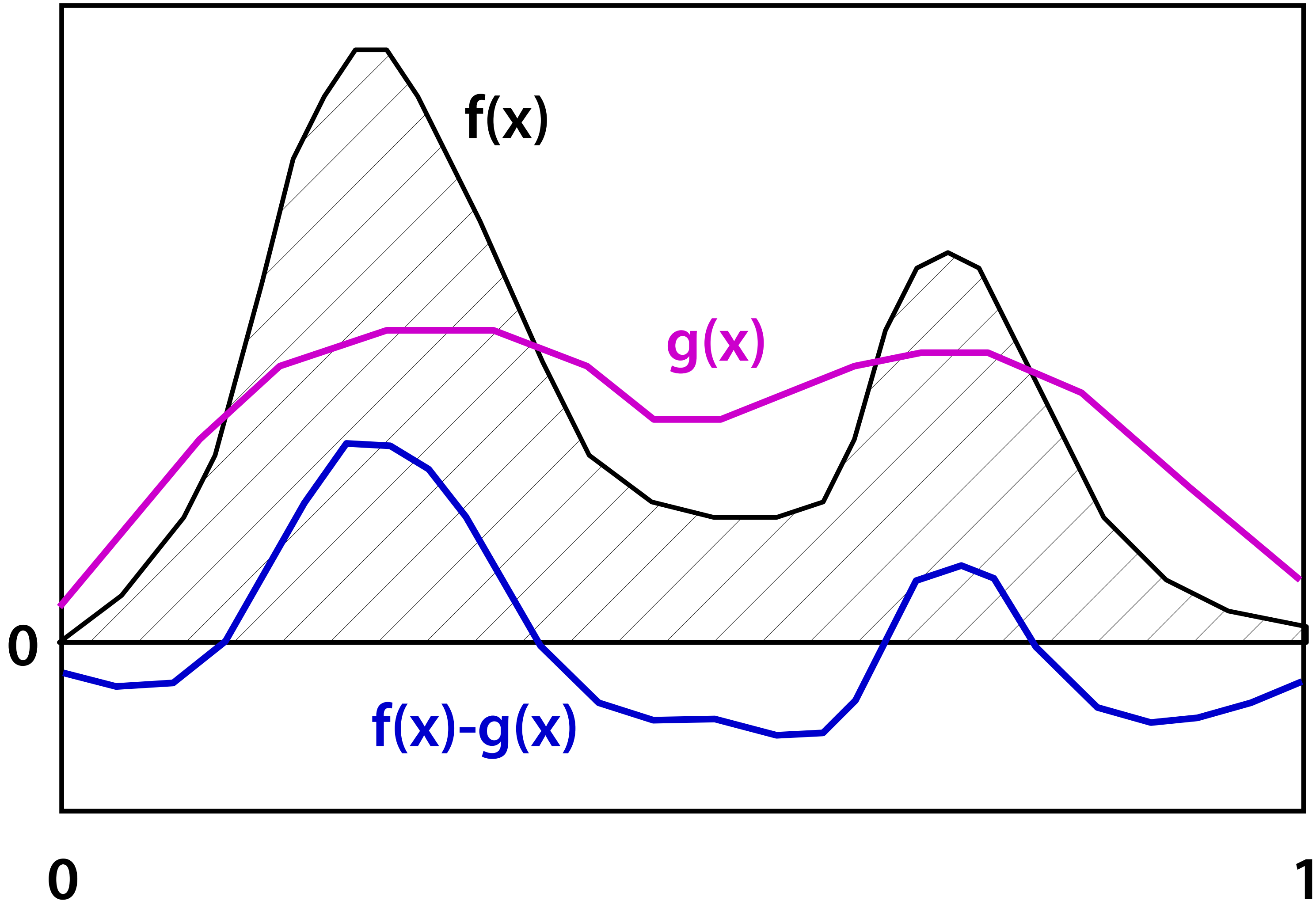
Power heuristics

β = **2**

Both samplings!



Control function



Control function

Function **g(x)** is **approximating the integrand**.
If we can integrate it easily (closed-form solution) …

$$I = \int_0^1 f(x)\,dx = \int_0^1 [f(x) - g(x)]\,dx + \underline{\int_0^1 g(x)\,dx} =$$

$$= \int_0^1 [f(x) - g(x)]\,dx + \underline{J} = \int_0^1 [f(x) - g(x) + J]\,dx$$

Unbiased estimate $\underline{\langle I \rangle_{con} = f(\xi) - g(\xi) + J}$

Part two – integral equations

Fredholm integral equation of the second kind

$$f(x) = g(x) + \int_0^1 K(x,y) \cdot f(y) \, dy$$

unknown

known functions

1. **Finite element methods** FEM (the whole function)
2. **Monte-Carlo methods** (local solution possible)

Recursive Monte-Carlo estimate

The right side is estimated **stochastically** (importance sampling) using PDFs $\mathbf{p_i(x)}$

$$\langle f(x) \rangle_r = g(x) + \frac{K(x,\xi_1)}{p_1(\xi_1)} \cdot \langle f(\xi_1) \rangle_r =$$

$$= g(x) + \frac{K(x,\xi_1)}{p_1(\xi_1)} \cdot \left[g(\xi_1) + \frac{K(\xi_1,\xi_2)}{p_2(\xi_2)} \cdot \langle f(\xi_2) \rangle_r \right]$$

$$= g(x) + \frac{K(x,\xi_1)}{p_1(\xi_1)} g(\xi_1) + \frac{K(x,\xi_1)}{p_1(\xi_1)} \frac{K(\xi_1,\xi_2)}{p_2(\xi_2)} g(\xi_2) + \ldots$$

Recursive Monte-Carlo estimate

$$\langle f(x) \rangle_r = \sum_{i=0}^{\infty} \left[\prod_{j=1}^{i} \frac{K(\xi_{j-1}, \xi_j)}{p_j(\xi_j)} \right] g(\xi_i), \qquad \xi_0 = x$$

$\{\xi_1, \xi_2, \xi_3, \dots\}$ is a **Markov chain**, if the probability **$p_i(x)$** is stationary – depends only on ξ_{i-1} (and not on **i**)

Functional formulation $\quad f = g + Tf$

Solution (Neumann series) $\quad f = g + Tg + T^2 g + \dots$

Russian roulette

While evaluation a **Neumann series** only finite number of terms can be considered

- **fixed cutoff** would introduce **bias** (reduction)

More correct random cutoff using **Russian roulette**

- estimate keeps **unbiased**

Can be applied to single integral estimate

- but is more useful for **infinite series cutoff**

Russian roulette for a single integral

Integral transformation

$$I = \int_0^1 f(x)\,dx = \int_0^P \frac{1}{P} f\left(\frac{t}{P}\right) dt \qquad 0 < P \le 1$$

Unbiased estimate with one random sample

$$\langle I \rangle_{Russ} = \begin{cases} \frac{1}{P} f\left(\frac{\xi}{P}\right) & \text{for} \quad \xi < P \\ 0 & \text{otherwise} \end{cases}$$

Russian roulette for integral equations

$$\left\langle f(x) \right\rangle_{Russ,r} = \sum_{i=0}^{k} \left[\prod_{j=1}^{i} \frac{K(\xi_{j-1}, \xi_j)}{P_j \cdot p_j(\xi_j)} \right] g(\xi_i), \qquad \xi_0 = x$$

$\{\xi_1, \xi_2, \ldots \xi_k\}$ is **finite** random walk (estimate $\langle \mathbf{f}(\xi_k) \rangle = \mathbf{0}$)

Every sample (step) ξ_i is chosen with **probability** $\mathbf{P}_i$ and from PDF (density) $\mathbf{p}_i(\mathbf{x})$

If random value $\tau_{i+1} > \mathbf{P}_{i+1}$, the whole process ends, otherwise a new step ξ_{i+1} is computed (new term)

Probabilities

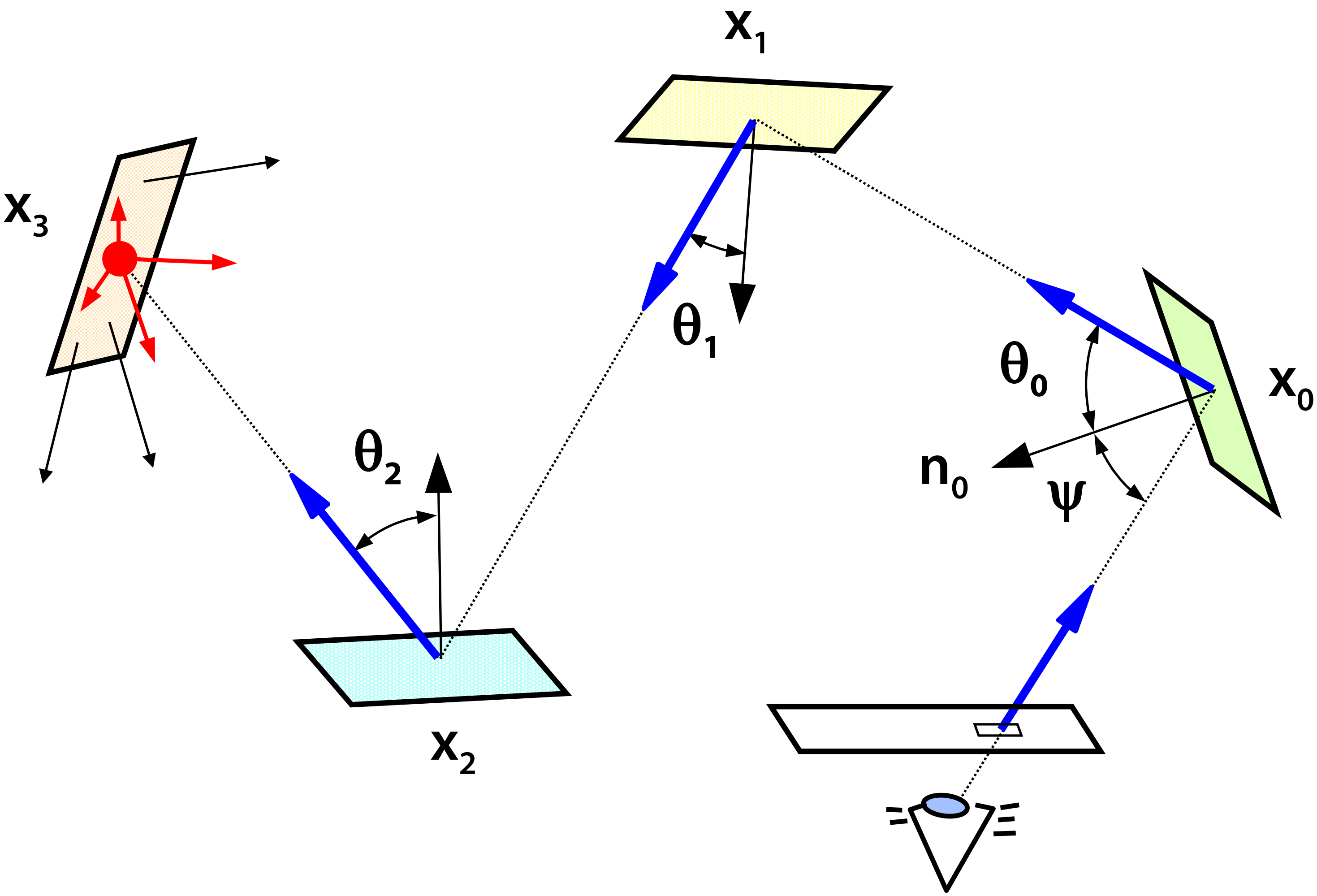
In physical contexts often $$\int_0^1 K(x, y)\, dy < 1$$

Then the kernel **K** can be used for construction of a **subcritical probability setup**

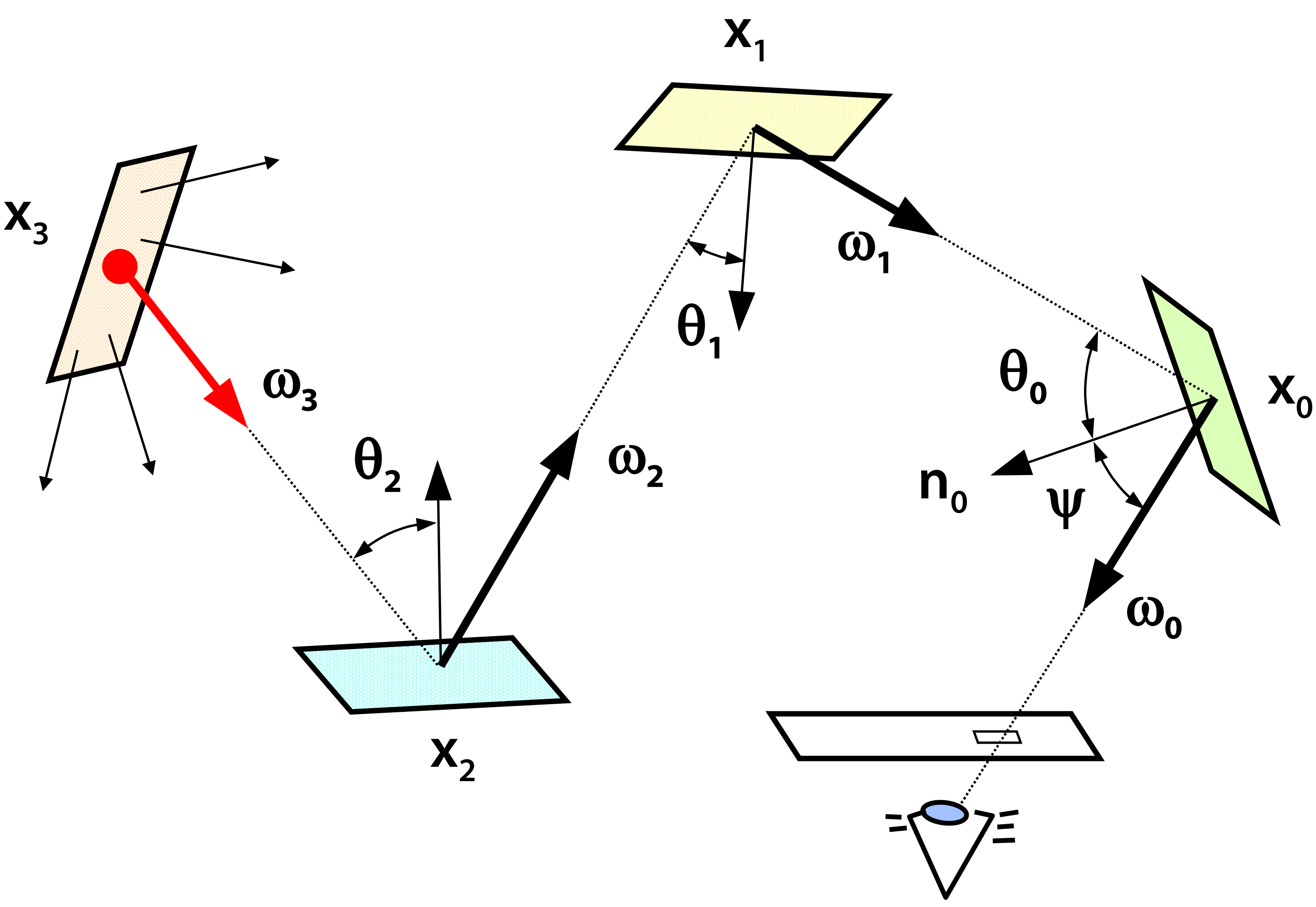
$$P_i = \int_0^1 K(\xi_{i-1}, y)\, dy, \qquad p_i(x) = \frac{K(\xi_{i-1}, x)}{P_i}$$

Estimate is reduced to $$\langle f(x) \rangle_{subcrit} = \sum_{i=1}^{k} g(\xi_i)$$

Path Tracing – walk from camera



Path Tracing – light propagation



Next event estimation (NEE)

Previous estimate uses to have **too high variance** (i.e. only small number of terms are nonzero)

A better way is to estimate **g(x)** using an "one step ahead" approach…

$$f(x) = g(x) + h(x)$$

$$h(x) = \int_0^1 K(x,y) \cdot f(y)\, dy =$$

$$= \underline{\int_0^1 K(x,y) \cdot g(y)\, dy} + \underline{\int_0^1 K(x,y) \cdot h(y)\, dy}$$

Next event estimation (NEE)

- **The 1st integral** can be estimated using PDF similar to **g(x) … no recursion!**
  - random value ζ_i from PDF $\mathbf{p_i(x)}$
- **The 2nd integral** (recursive) is estimated using importance sampling with subcritical probability (kernel **K**)
  - random value ξ_i from PDF $\mathbf{K(\xi_{i-1},x)/P_i}$

$$\langle h(x) \rangle_{nextev} = \frac{K(x,\zeta_1)\, g(\zeta_1)}{p_1(\zeta_1)} + \langle h(\xi_1) \rangle_{nextev}$$

Next event estimation

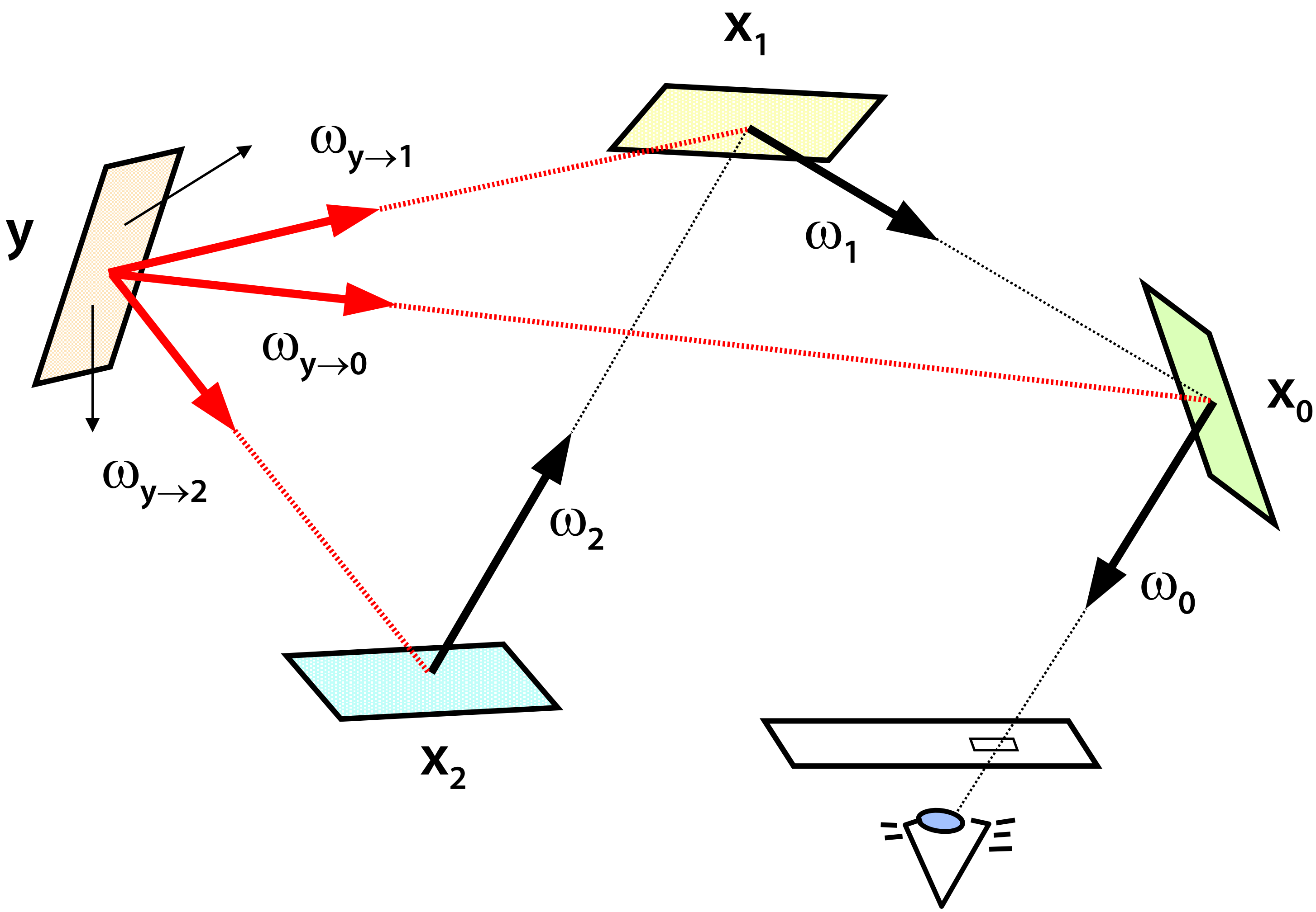
Estimation of the function **h**

$$\langle \mathbf{h}(\mathbf{x}) \rangle_{\mathbf{nextev}} = \sum_{i=1}^{k} \frac{\mathbf{K}(\xi_{i-1}, \zeta_i)\, \mathbf{g}(\zeta_i)}{\mathbf{p}_i(\zeta_i)}$$

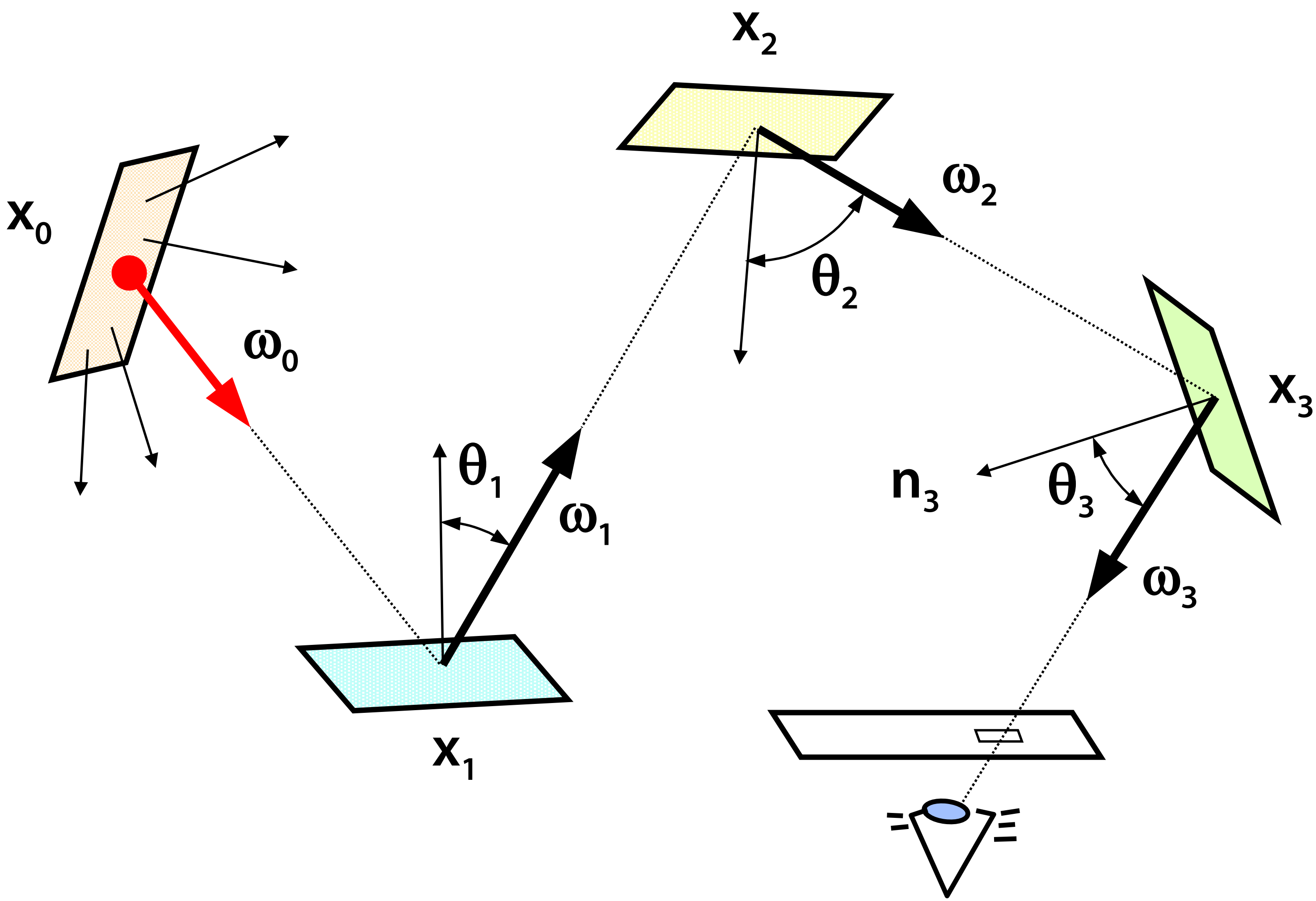
Integral NEE estimate

$$\langle \mathbf{f}(\mathbf{x}) \rangle_{\mathbf{nextev}} = \mathbf{g}(\mathbf{x}) + \sum_{i=1}^{k} \frac{\mathbf{K}(\xi_{i-1}, \zeta_i)\, \mathbf{g}(\zeta_i)}{\mathbf{p}_i(\zeta_i)}$$

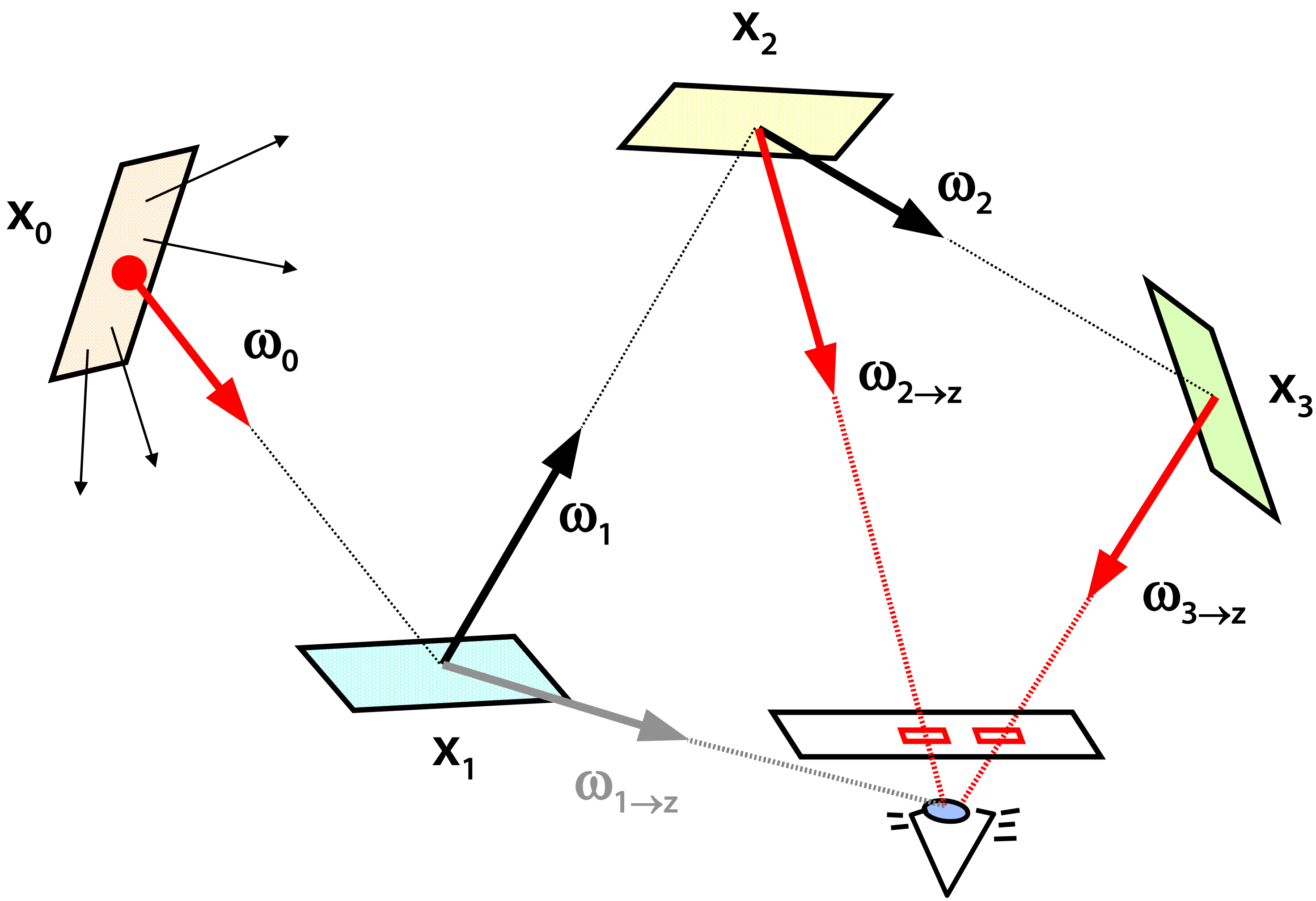
NEE for Path Tracing



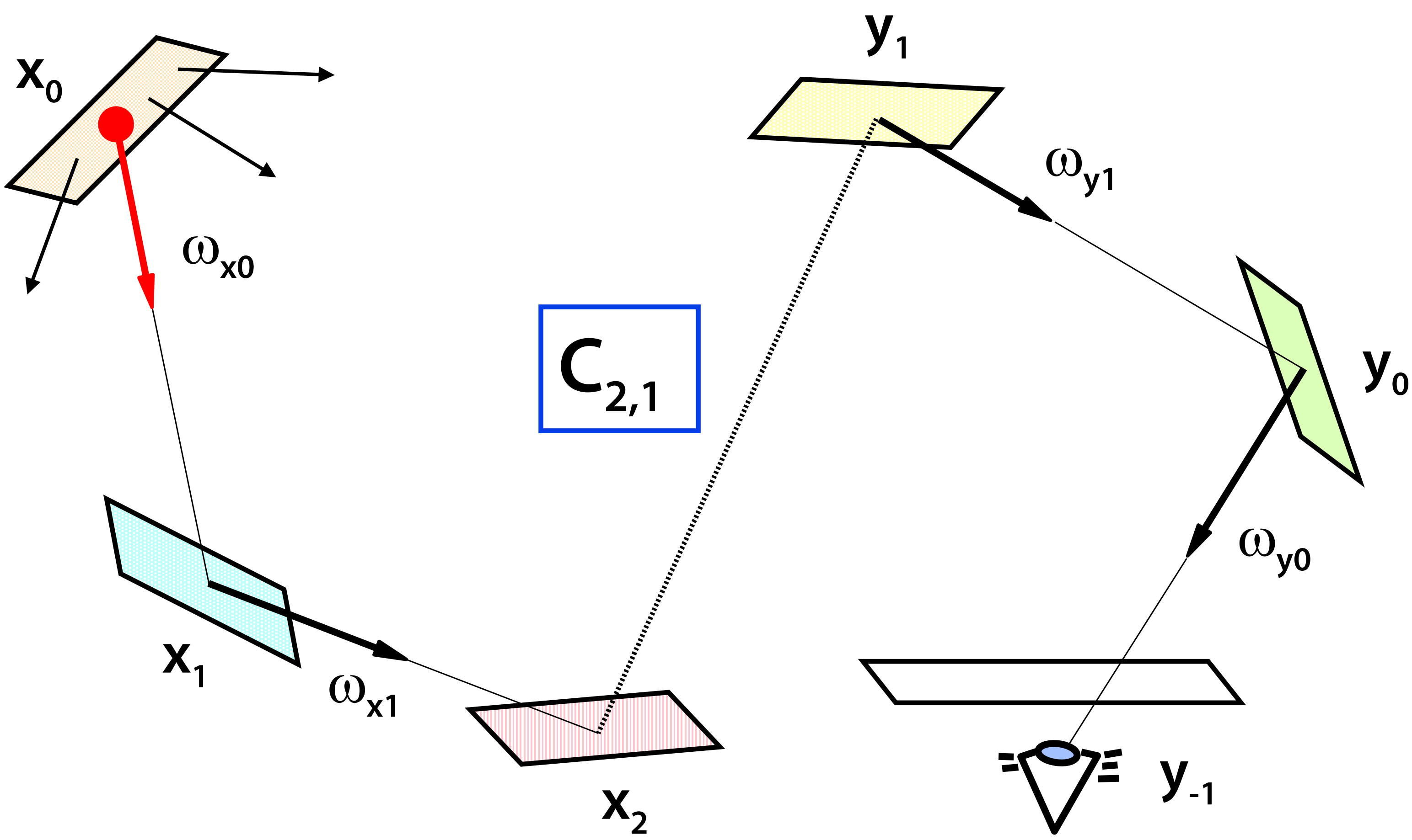
Light Tracing – light propagation



NEE for Light Tracing



Bidirectional Path Tracing



Bidirectional Path Tracing example

© 1995 Eric Veach,
Leonidas J. Guibas

References

E. Lafortune: ***Mathematical Models and Monte Carlo Algorithms for Physically Based Rendering***, PhD thesis, KU Leuven, 29-63

M. Kalos, P. Whitlock: ***Monte Carlo Methods***, John Wiley & Sons, 1986, 89-116

A. Glassner: ***Principles of Digital Image Synthesis***, Morgan Kaufmann, 1995, 840-864