Physically Based Rendering (600.657)

Monte Carlo Integration

Probability

<u>Definition</u>:

The *probability density function* (PDF) $p(x)$ of a random variable X in a domain Ω describes the relative likelihood for this variable to occur at a given point in Ω:

1. $0 \leq p(x)$
2. $\int_\Omega p(x)dx = 1$

Probability

Note:

Given any subdomain $D \subset \Omega$, the probability of the variable occurring within D is:

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

Probability

Definition:

For a (1D) domain Ω=[*a*,*b*] the *cumulative distribution function* (CDF) *P*(*x*) of a random variable *X* is the probability that a value from the variable's distribution is less than or equal to so some value *x*:

$$P(x) = \Pr\{X \leq x\} = \int_a^x p(x')dx'$$

Probability

Note:

Given a CDF on the domain Ω=[*a*,*b*], we have:

$$P(a) = \int_a^a p(x)dx = 0 \qquad P(b) = \int_a^b p(x)dx = 1$$

Probability

Note:

Given a CDF on the domain Ω=[*a*,*b*], we have:

$$P(a)=\int_a^a p(x)dx=0 \qquad P(b)=\int_a^b p(x)dx=1$$

And more generally:

$$p(x)=\left.\frac{dP}{dx}\right|_x$$

Probability

<u>Definition</u>:

Given a PDF $p(x_1,x_2)$ on the domain $\Omega_1\times\Omega_2$, the *marginal density function* of $x_1\in \Omega_1$ is obtained by integrating out one of the dimensions:

$$p(x_1) = \int_{\Omega_2} p(x_1, x_2)dx_2$$

Probability

Note:

The marginal density function is also a PDF because it is non-negative and:

$$\int_{\Omega_1} p(x_1)dx_1 = \int_{\Omega_1}\int_{\Omega_2} p(x_1, x_2)dx_2dx_1 = 1$$

Probability

Definition:

The marginal densities p_1 and p_2 are *independent* if the probability distribution on $\Omega_1 \times \Omega_2$ is the product of marginal probability distributions on $\Omega_1$and Ω_2:

$$p(x_1, x_2) = p(x_1)\, p(x_2)$$

Probability

Definition:

Given a PDF $p(x_1,x_2)$ on the domain $\Omega_1 \times \Omega_2$, the *conditional density function* of $x_2 \in \Omega_2$ given $x_1 \in \Omega_1$ is obtained by integrating out one of the dimensions:

$$p(x_2 \mid x_1) = \frac{p(x_1, x_2)}{p(x_1)}$$

Probability

Note:

The conditional density function is a PDF because it is non-negative and:

$$\int_{\Omega_2} p(x_2 \mid x_1) dx_2 = \int_{\Omega_2} \frac{p(x_1, x_2)}{p(x_1)} dx_2 = \frac{1}{p(x_1)} \int_{\Omega_2} p(x_1, x_2) dx_2 = \frac{p(x_1)}{p(x_1)} = 1$$

Probability

Definition:

Given a PDF *p* on a domain Ω, the *expected value,* $E_p[f]$, of a function *f* is the average value of the function over a distribution of values *p*(*x*) over its domain:

$$E_p\left[f\right]=\int_{\Omega} f(x)\,p(x)\,dx$$

Probability

Note:

For functions f, f_1, and f_2 and constant c:

$$E_p[c] = \int cp(x)dx = c\int p(x)dx = c$$

Probability

Note:

For functions f, f_1, and f_2 and constant c:

$$E_p\left[c\right] = \int cp(x)dx = c\int p(x)dx = c$$

$$E_p\left[cf\right] = \int cf(x)p(x)dx = c\int f(x)p(x)dx = cE_p[f]$$

Probability

Note:

For functions f, f_1, and f_2 and constant c:

$$E_p[c] = \int c p(x) dx = c \int p(x) dx = c$$

$$E_p[cf] = \int cf(x) p(x) dx = c \int f(x) p(x) dx = cE_p[f]$$

$$E_p[f_1 + f_2] = \int \left(f_1(x) + f_2(x) \right) p(x) dx = \int f_1(x) p(x) dx + \int f_2(x) p(x) dx$$
$$= E_p[f_1] + E_p[f_2]$$

Probability

Note:

Given a probability distribution p on $\Omega_1 \times \Omega_2$, and given function f_1 on Ω_1 and f_2 on Ω_2:

$$E_p\left[f_1 + f_2\right] = \iint_{\Omega_1 \times \Omega_2} \left(f_1(x_1) + f_2(x_2)\right) p(x_1, x_2)\, dx_2\, dx_1$$

Probability

Note:

Given a probability distribution p on $\Omega_1 \times \Omega_2$, and given function f_1 on Ω_1 and f_2 on Ω_2:

$$E_p\left[f_1 + f_2\right] = \iint_{\Omega_1\times\Omega_2} \left(f_1(x_1) + f_2(x_2)\right) p(x_1, x_2)\, dx_2\, dx_1$$

$$= \int_{\Omega_1} f_1(x_1) \int_{\Omega_2} p(x_1, x_2)\, dx_2\, dx_1 + \int_{\Omega_2} f(x_2) \int_{\Omega_1} p(x_1, x_2)\, dx_1\, dx_2$$

Probability

Note:

Given a probability distribution *p* on $\Omega_1 \times \Omega_2$, and given function f_1 on Ω_1 and f_2 on Ω_2:

$$E_p\left[f_1 + f_2\right] = \iint_{\Omega_1 \times \Omega_2} \left(f_1(x_1) + f_2(x_2)\right) p(x_1, x_2)\, dx_2\, dx_1$$

$$= \int_{\Omega_1} f_1(x_1) \int_{\Omega_2} p(x_1, x_2)\, dx_2\, dx_1 + \int_{\Omega_2} f(x_2) \int_{\Omega_1} p(x_1, x_2)\, dx_1\, dx_2$$

$$= \int_{\Omega} f(x_1)\, p_1(x_1)\, dx_1 + \int_{\Omega} f(x_2)\, p_2(x_2)\, dx_2$$

Probability

Note:

Given a probability distribution p on $\Omega_1 \times \Omega_2$, and given function f_1 on Ω_1 and f_2 on Ω_2:

$$
\begin{aligned}
E_p\left[f_1 + f_2\right] &= \iint_{\Omega_1 \times \Omega_2} \big(f_1(x_1) + f_2(x_2)\big) p(x_1, x_2)\, dx_2\, dx_1 \\
&= \int_{\Omega_1} f_1(x_1) \int_{\Omega_2} p(x_1, x_2)\, dx_2\, dx_1 + \int_{\Omega_2} f(x_2) \int_{\Omega_1} p(x_1, x_2)\, dx_1\, dx_2 \\
&= \int_{\Omega} f(x_1)\, p_1(x_1)\, dx_1 + \int_{\Omega} f(x_2)\, p_2(x_2)\, dx_2 \\
&= E_{p_1}\left[f_1\right] + E_{p_2}\left[f_2\right]
\end{aligned}
$$

where p_1 and p_2 are the marginal distributions.

Probability

Note:

Given independent PDFs p_1 on Ω_1 and p_2 on Ω_2, and given functions f_1 on Ω_1 and f_2 on Ω_2:

$$E_p\left[f_1(x_1)\cdot f_2(x_2)\right] = \iint_{\Omega_1\times\Omega_2} f_1(x_1)\, f_2(x_2)\, p(x_1, x_2)\, dx_2\, dx_1$$

Probability

Note:

Given independent PDFs p_1 on Ω_1 and p_2 on Ω_2, and given functions f_1 on Ω_1 and f_2 on Ω_2:

$$\begin{aligned} E_p\left[f_1(x_1)\cdot f_2(x_2)\right] &= \iint\limits_{\Omega_1\times\Omega_2} f_1(x_1)\, f_2(x_2)\, p(x_1,x_2)\,dx_2\,dx_1 \\ &= \iint\limits_{\Omega_1\times\Omega_2} f_1(x_1)\, p_1(x_1)\,dx_1\, f_2(x_2)\, p_2(x_2)\,dx_2 \end{aligned}$$

Probability

Note:

Given independent PDFs p_1 on Ω_1 and p_2 on Ω_2, and given functions f_1 on Ω_1 and f_2 on Ω_2:

$$
\begin{aligned}
E_p\left[f_1(x_1)\cdot f_2(x_2)\right] &= \iint_{\Omega_1\times\Omega_2} f_1(x_1)\, f_2(x_2)\, p(x_1, x_2)\, dx_2\, dx_1 \\
&= \iint_{\Omega_1\times\Omega_2} f_1(x_1)\, p_1(x_1)\, dx_1\, f_2(x_2)\, p_2(x_2)\, dx_2 \\
&= \int_{\Omega_1} f_1(x_1)\, p_1(x_1)\, dx_1 \int_{\Omega_2} f_2(x_2)\, p_2(x_2)\, dx_2
\end{aligned}
$$

Probability

Note:

Given independent PDFs p_1 on Ω_1 and p_2 on Ω_2, and given functions f_1 on Ω_1 and f_2 on Ω_2:

$$
\begin{aligned}
E_p\left[f_1(x_1)\cdot f_2(x_2)\right] &= \iint_{\Omega_1\times\Omega_2} f_1(x_1)\,f_2(x_2)\,p(x_1,x_2)\,dx_2\,dx_1 \\
&= \iint_{\Omega_1\times\Omega_2} f_1(x_1)\,p_1(x_1)\,dx_1\,f_2(x_2)\,p_2(x_2)\,dx_2 \\
&= \int_{\Omega_1} f_1(x_1)\,p_1(x_1)\,dx_1 \int_{\Omega_2} f_2(x_2)\,p_2(x_2)\,dx_2 \\
&= E_{p_1}\left[f_1\right]\cdot E_{p_2}\left[f_2\right]
\end{aligned}
$$

Probability

Definition:

The *variance*, $V_p[f]$, of a function f is the expected deviation of the function from its expected value:

$$V_p[f] = E_p\left[\left(f - E_p[f]\right)^2\right]$$

Probability

Definition:

The *variance, $V_p[f]$*, of a function f is the expected deviation of the function from its expected value:

$$V_p[f] = E_p\left[\left(f - E_p[f]\right)^2\right]$$
$$= E_p\left[f^2 + E_p[f]^2 - 2E_p[f]f\right]$$

Probability

Definition:

The *variance,* $V_p[f]$, of a function f is the expected deviation of the function from its expected value:

$$
\begin{aligned}
V_p[f] &= E_p\left[\left(f - E_p[f]\right)^2\right] \\
&= E_p\left[f^2 + E_p[f]^2 - 2E_p[f]f\right] \\
&= E_p\left[f^2\right] + E_p[f]^2 - 2E_p\left[E_p[f]f\right]
\end{aligned}
$$

Probability

Definition:

The *variance, $V_p[f]$,* of a function f is the expected deviation of the function from its expected value:

$$
\begin{aligned}
V_p[f] &= E_p\left[\left(f - E_p[f]\right)^2\right] \\
&= E_p\left[f^2 + E_p[f]^2 - 2E_p[f]f\right] \\
&= E_p\left[f^2\right] + E_p\left[f\right]^2 - 2E_p\left[E_p[f]f\right] \\
&= E_p\left[f^2\right] + E_p\left[f\right]^2 - 2E_p\left[f\right]E_p[f]
\end{aligned}
$$

Probability

<u>Definition</u>:

The *variance, $V_p[f]$*, of a function *f* is the expected deviation of the function from its expected value:

$$
\begin{aligned}
V_p[f] &= E_p\left[(f - E_p[f])^2\right] \\
&= E_p\left[f^2 + E_p[f]^2 - 2E_p[f]f\right] \\
&= E_p\left[f^2\right] + E_p[f]^2 - 2E_p\left[E_p[f]f\right] \\
&= E_p\left[f^2\right] + E_p[f]^2 - 2E_p[f]E_p[f] \\
&= E_p\left[f^2\right] - E_p[f]^2
\end{aligned}
$$

Probability

Note:

For functions *f* and constant *c*:

$$V_p[cf] = E_p\left[(cf)^2\right] - E_p[cf]^2 = c^2\left(E_p\left[f^2\right] - E_p[f]^2\right) = c^2 V_p[f]$$

Probability

Note:

Given independent PDFs p_1 on Ω_1 and p_2 on Ω_2, and given functions f_1 on Ω_1 and f_2 on Ω_2:

$$V_p[f_1 + f_2] = E_p\left[(f_1 + f_2)^2\right] - E_p[f_1 + f_2]^2$$

Probability

Note:

Given independent PDFs p_1 on Ω_1 and p_2 on Ω_2, and given functions f_1 on Ω_1 and f_2 on Ω_2:

$$
\begin{aligned}
V_p\left[f_1 + f_2\right] &= E_p\left[\left(f_1 + f_2\right)^2\right] - E_p\left[f_1 + f_2\right]^2 \\
&= E_p\left[f_1^2 + f_2^2 + 2 f_1 f_2\right] - E_p\left[f_1 + f_2\right]^2
\end{aligned}
$$

Probability

Note:

Given independent PDFs p_1 on Ω_1 and p_2 on Ω_2, and given functions f_1 on Ω_1 and f_2 on Ω_2:

$$
\begin{aligned}
V_p[f_1 + f_2] &= E_p\left[(f_1 + f_2)^2\right] - E_p[f_1 + f_2]^2 \\
&= E_p\left[f_1^2 + f_2^2 + 2 f_1 f_2\right] - E_p[f_1 + f_2]^2 \\
&= E_{p_1}\left[f_1^2\right] + E_{p_2}\left[f_2^2\right] + 2E_{p_1}[f_1]E_{p_2}[f_2] - \left(E_{p_1}[f_1] + E_{p_2}[f_2]\right)^2
\end{aligned}
$$

Probability

Note:

Given independent PDFs p_1 on Ω_1 and p_2 on Ω_2, and given functions f_1 on Ω_1 and f_2 on Ω_2:

$$
\begin{aligned}
V_p[f_1+f_2] &= E_p\left[(f_1+f_2)^2\right]-E_p[f_1+f_2]^2 \\
&= E_p\left[f_1^2+f_2^2+2f_1f_2\right]-E_p[f_1+f_2]^2 \\
&= E_{p_1}\left[f_1^2\right]+E_{p_2}\left[f_2^2\right]+2E_{p_1}[f_1]E_{p_2}[f_2]-\left(E_{p_1}[f_1]+E_{p_2}[f_2]\right)^2 \\
&= E_{p_1}\left[f_1^2\right]-E_{p_1}[f_1]^2+E_{p_2}\left[f_2^2\right]-E_{p_2}[f_2]^2
\end{aligned}
$$

Probability

<u>Note</u>:

Given independent PDFs p_1 on Ω_1 and p_2 on Ω_2, and given functions f_1 on Ω_1 and f_2 on Ω_2:

$$
\begin{aligned}
V_p\left[f_1+f_2\right] &= E_p\left[\left(f_1+f_2\right)^2\right]-E_p\left[f_1+f_2\right]^2 \\
&= E_p\left[f_1^2+f_2^2+2f_1f_2\right]-E_p\left[f_1+f_2\right]^2 \\
&= E_{p_1}\left[f_1^2\right]+E_{p_2}\left[f_2^2\right]+2E_{p_1}\left[f_1\right]E_{p_2}\left[f_2\right]-\left(E_{p_1}\left[f_1\right]+E_{p_2}\left[f_2\right]\right)^2 \\
&= E_{p_1}\left[f_1^2\right]-E_{p_1}\left[f_1\right]^2+E_{p_2}\left[f_2^2\right]-E_{p_2}\left[f_2\right]^2 \\
&= V_{p_1}\left[f_1\right]+V_{p_2}\left[f_2\right]
\end{aligned}
$$

Probability

Definition:

Given a PDF $p(x_1,x_2)$ on the domain $\Omega_1\times\Omega_2$ and given a function f on $\Omega_1\times\Omega_2$, the *conditional expectation* over Ω_2 given $x_1\in\Omega_1$ is:

$$E_{\Omega_1}\left[f \mid x_1\right]=\int_{\Omega_2} f(x_1,x_2)\cdot p(x_2 \mid x_1)dx_2$$

Probability

Definition:

Given a PDF $p(x_1,x_2)$ on the domain $\Omega_1\times\Omega_2$ and given a function f on $\Omega_1\times\Omega_2$, the *conditional expectation* over Ω_2 given $x_1\in\Omega_1$ is:

$$E_{\Omega_1}\left[f \mid x_1\right]=\int_{\Omega_2} f(x_1,x_2)\cdot p(x_2 \mid x_1)dx_2$$

Similarly, the conditional variance over Ω_2 given $x_1\in\Omega_1$ is:

$$V_{\Omega_1}\left[f \mid x_1\right]=\int_{\Omega_2}\left(f(x_1,x_2)-E[f \mid x_1]\right)^2\cdot p(x_2 \mid x_1)dx_2$$

Probability

$$E_{\Omega_1}\left[f \mid x_1\right] = \int_{\Omega_2} f(x_1, x_2) \cdot p(x_2 \mid x_1) dx_2$$

$$V_{\Omega_1}\left[f \mid x_1\right] = \int_{\Omega_2} \left(f(x_1, x_2) - E[f \mid x_1]\right)^2 \cdot p(x_2 \mid x_1) dx_2$$

<u>Note</u>:

$$V_{\Omega_1 \times \Omega_1}[f] = E_{\Omega_1 \times \Omega_1}[f^2] - \left(E_{\Omega_1 \times \Omega_1}[f]\right)^2$$

Probability

$$E_{\Omega_1}\left[f \mid x_1\right] = \int_{\Omega_2} f(x_1, x_2) \cdot p(x_2 \mid x_1) dx_2$$

$$V_{\Omega_1}\left[f \mid x_1\right] = \int_{\Omega_2} \left(f(x_1, x_2) - E[f \mid x_1]\right)^2 \cdot p(x_2 \mid x_1) dx_2$$

<u>Note</u>:

$$\begin{aligned} V_{\Omega_1 \times \Omega_1}[f] &= E_{\Omega_1 \times \Omega_1}[f^2] - \left(E_{\Omega_1 \times \Omega_1}[f]\right)^2 \\ &= E_{\Omega_1}\left[E_{\Omega_2}[f^2 \mid x_1]\right] - \left(E_{\Omega_1}\left[E_{\Omega_2}[f \mid x_1]\right]\right)^2 \end{aligned}$$

Probability

$$E_{\Omega_1}\left[f \mid x_1\right] = \int_{\Omega_2} f(x_1, x_2) \cdot p(x_2 \mid x_1) dx_2$$

$$V_{\Omega_1}\left[f \mid x_1\right] = \int_{\Omega_2} \left(f(x_1, x_2) - E[f \mid x_1] \right)^2 \cdot p(x_2 \mid x_1) dx_2$$

<u>Note</u>:

$$\begin{aligned} V_{\Omega_1 \times \Omega_1}[f] &= E_{\Omega_1 \times \Omega_1}[f^2] - \left(E_{\Omega_1 \times \Omega_1}[f] \right)^2 \\ &= E_{\Omega_1}\left[E_{\Omega_2}[f^2 \mid x_1] \right] - \left(E_{\Omega_1}\left[E_{\Omega_2}[f \mid x_1] \right] \right)^2 \\ &= E_{\Omega_1}\left[E_{\Omega_2}[f^2 \mid x_1] \right] - E_{\Omega_1}\left[\left(E_{\Omega_2}[f \mid x_1] \right)^2 \right] + \\ &\quad E_{\Omega_1}\left[\left(E_{\Omega_2}[f \mid x_1] \right)^2 \right] - \left(E_{\Omega_1}\left[E_{\Omega_2}[f \mid x_1] \right] \right)^2 \end{aligned}$$

Probability

$$E_{\Omega_1}\left[f \mid x_1\right] = \int_{\Omega_2} f(x_1, x_2) \cdot p(x_2 \mid x_1)\, dx_2$$

$$V_{\Omega_1}\left[f \mid x_1\right] = \int_{\Omega_2} \left(f(x_1, x_2) - E[f \mid x_1]\right)^2 \cdot p(x_2 \mid x_1)\, dx_2$$

Note:

$$\begin{aligned}
V_{\Omega_1\times\Omega_1}[f] &= E_{\Omega_1\times\Omega_1}[f^2] - \left(E_{\Omega_1\times\Omega_1}[f]\right)^2 \\
&= E_{\Omega_1}\left[E_{\Omega_2}[f^2 \mid x_1]\right] - \left(E_{\Omega_1}\left[E_{\Omega_2}[f \mid x_1]\right]\right)^2 \\
&= E_{\Omega_1}\left[E_{\Omega_2}[f^2 \mid x_1]\right] - E_{\Omega_1}\left[\left(E_{\Omega_2}[f \mid x_1]\right)^2\right] + \\
&\quad E_{\Omega_1}\left[\left(E_{\Omega_2}[f \mid x_1]\right)^2\right] - \left(E_{\Omega_1}\left[E_{\Omega_2}[f \mid x_1]\right]\right)^2 \\
&= E_{\Omega_1}\left[V_{\Omega_2}[f \mid x_1]\right] + V_{\Omega_1}\left[E_{\Omega_2}[f \mid x_1]\right]
\end{aligned}$$

Monte Carlo Estimator

Definition:

Given a PDF p on domain Ω corresponding to random variable X, and given a function f on Ω, the (n-th) Monte Carlo estimate of the integral of f over Ω is:

$$I_n(X) = \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{p(X_i)}$$

Monte Carlo Estimator

Note:

The expected value of the estimate is:

$$E_p\left[I_n\right] = E_p\left[\frac{1}{n}\sum_{i=1}^{n}\frac{f}{p}\right] = \frac{1}{n}\sum_{i=1}^{n}E_p\left[\frac{f}{p}\right] = \frac{1}{n}\sum_{i=1}^{n}\int_{\Omega}\frac{f(x)}{p(x)}\,p(x)dx = \int_{\Omega} f(x)dx$$

Monte Carlo Estimator

Note:

The expected value of the estimate is:

$$E_p\left[I_n\right] = \int_{\Omega} f(x)dx$$

The expected value of the estimate is the integral of the function, regardless of the PDF.
(So long as $p(x) \neq 0$ when $f(x) \neq 0$.)

Monte Carlo Estimator

Note:

The expected value of the estimate is:

$$E_p\left[I_n\right] = \int_{\Omega} f(x)dx$$

The variance of the estimate is:

$$V_p\left[I_n\right] = V_p\left[\frac{1}{n}\sum_{i=1}^{n}\frac{f}{p}\right] = \frac{1}{n^2}\sum_{i=1}^{n}V_p\left[\frac{f}{p}\right] = \frac{1}{n}V_p\left[\frac{f}{p}\right]$$

Monte Carlo Estimator

Note:

The expected value of the estimate is:

$$E_p\left[I_n\right] = \int_{\Omega} f(x)dx$$

The variance of the estimate is:

$$V_p\left[I_n\right] = \frac{1}{n} V_p\left[\frac{f}{p}\right]$$

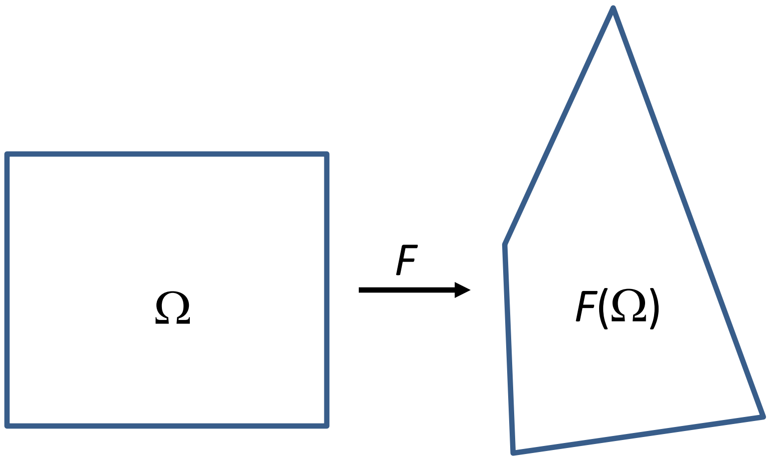
Variance decreases as:

1. We use more samples
2. The variance of *f/p* decreases (e.g. if $p \propto f$.)

Sampling

<u>Challenge</u>:

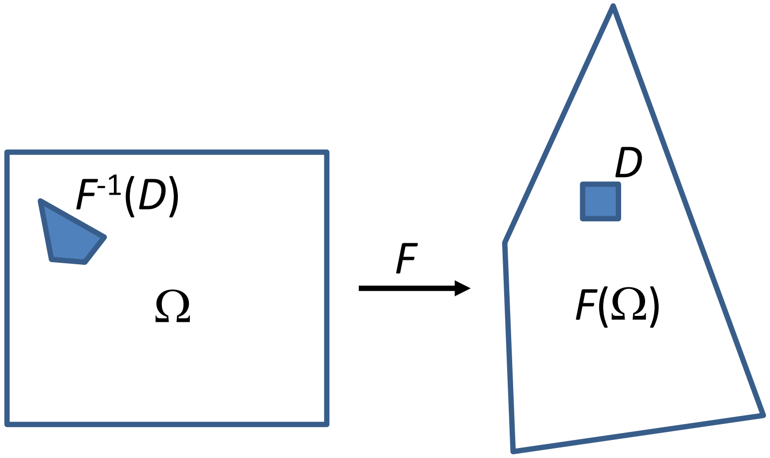
Given a random variable X on domain Ω with PDF p, and given a bijective map $F:\Omega\rightarrow F(\Omega)$ what is the PDF q of the random variable $F(X)$?



Sampling

<u>Challenge</u>:

For a given domain $D \subset F(\Omega)$, we know that the probability that $F(x)$ is in D is equal to the probability that x in $F^{-1}(D)$.



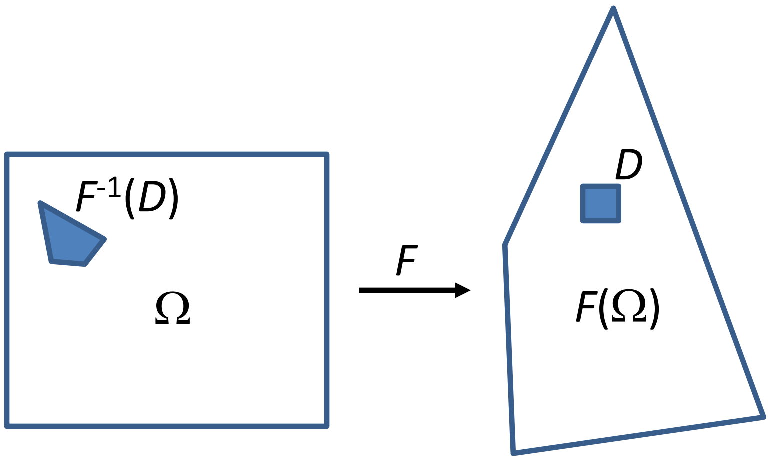
Sampling

<u>Challenge</u>:

For a given domain $D \subset F(\Omega)$, we know that the probability that $F(x)$ is in D is equal to the probability that x in $F^{-1}(D)$.

Thus, the differential probability of choosing the point $F(x)$ is:

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$



$$q\left(F(x)\right)=\frac{p(x)}{\left|J_F(x)\right|}$$

Sampling

<u>Definition</u>:

A *canonical uniform random variable* ξ is a random variable that takes on all values in the domain Ω=[0,1] with equal probability:

$$p(x)=\begin{cases}1 & x\in[0,1]\\ 0 & \text{otherwise}\end{cases}$$

$$q\big(F(x)\big) = \frac{p(x)}{\left|J_F(x)\right|}$$

Sampling

<u>Challenge</u>:

Given a canonical uniform random variable ξ, how do we generate a uniform random variable in the range $[a,b]$?

Sampling

$$q\left(F(x)\right)=\frac{p(x)}{\left|J_F(x)\right|}$$

<u>Solution</u>:

We would like a function $F:[0,1]\rightarrow[a,b]$ such that the probability of choosing $F(\xi)$ is $1/(b-a)$.

$$q\left(F(x)\right)=\frac{p(x)}{\left|J_F(x)\right|}$$

Sampling

<u>Solution</u>:

We would like a function F:[0,1]$\rightarrow$[a,b] such that the probability of choosing $F(\xi)$ is $1/(b-a)$.

Thus, we get:

$$\begin{cases} \dfrac{1}{b-a} & F(x)\in[a,b] \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \dfrac{1}{\left|J_F(x)\right|} & x\in[0,1] \\ 0 & \text{otherwise} \end{cases}$$

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

Sampling

Solution:

We would like a function F:[0,1]→[a,b] such that the probability of choosing $F(\xi)$ is 1/(b-a).

Thus, we get:

$$\begin{cases} \dfrac{1}{b-a} & F(x) \in [a,b] \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \dfrac{1}{\left|J_F(x)\right|} & x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

or in other words:

$$F'(x) = \pm(b-a) \quad \forall x \in [0,1]$$

Sampling

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

<u>Solution</u>:

We would like a function *F*:[0,1]→[*a*,*b*] such that the probability of choosing *F*(ξ) is 1/(*b*-*a*).

Thus, we get:

$$\begin{cases} \dfrac{1}{b-a} & F(x) \in [a,b] \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \dfrac{1}{\left|J_F(x)\right|} & x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

or in other words:

$$F'(x) = \pm(b-a) \quad \forall x \in [0,1] \;\Rightarrow\; F(x) = \pm\left(b-a\right)x + c \quad \forall x \in [0,1]$$

Sampling

$$q\left(F(x)\right)=\frac{p(x)}{\left|J_F(x)\right|}$$

<u>Solution</u>:

We would like a function F:[0,1]→[a,b] such that the probability of choosing $F(\xi)$ is 1/(b-a).

Thus, we get:

$$\begin{cases}\dfrac{1}{b-a} & F(x)\in[a,b]\\ 0 & \text{otherwise}\end{cases}=\begin{cases}\dfrac{1}{\left|J_F(x)\right|} & x\in[0,1]\\ 0 & \text{otherwise}\end{cases}$$

or in other words:

$$F'(x)=\pm(b-a)\quad\forall x\in[0,1]\;\Rightarrow\; F(x)=\pm\left(b-a\right)x+c\quad\forall x\in[0,1]$$

$$\Rightarrow\; F(x)=\begin{Bmatrix}(b-a)x+a\\(a-b)x+b\end{Bmatrix}\quad\forall x\in[0,1]$$

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

Sampling

Challenge:

How about if we would like to generate a random variable in the range [*a*,*b*] with PDF *q*?

$$q\big(F(x)\big) = \frac{p(x)}{\left|J_F(x)\right|}$$

Sampling

<u>Recall</u>:

Given *F*, the derivative of the inverse of *F* is:

$$\left(F^{-1}\right)' = \frac{1}{F' \circ F^{-1}}$$

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

Sampling

<u>Recall</u>:

Given *F*, the derivative of the inverse of *F* is:

$$\left(F^{-1}\right)' = \frac{1}{F' \circ F^{-1}}$$

This follows by the chain-rule:

$$\left(F \circ F^{-1}\right)' = 1$$

$$\Downarrow$$

$$\left(F' \circ F^{-1}\right) \cdot \left(F^{-1}\right)' = 1$$

$$\Downarrow$$

$$\left(F^{-1}\right)' = \frac{1}{\left(F' \circ F^{-1}\right)}.$$

Sampling

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

$$\left(F^{-1}\right)' = \frac{1}{F' \circ F^{-1}}$$

<u>Challenge</u>:

How about if we would like to generate a random variable in the range [*a*,*b*] with PDF *q*?

<u>Solution</u>:

In this case we get:

$$F' = \pm \frac{1}{q \circ F}$$

Sampling

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

$$\left(F^{-1}\right)' = \frac{1}{F' \circ F^{-1}}$$

Challenge:

How about if we would like to generate a random variable in the range [*a*,*b*] with PDF *q*?

Solution:

In this case we get:

$$F' = \pm \frac{1}{q \circ F}$$

Integrating to get $Q'(x)=q(x)$:

$$F' = \pm \frac{1}{Q' \circ F}$$

Sampling

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

$$\left(F^{-1}\right)' = \frac{1}{F' \circ F^{-1}}$$

Challenge:

How about if we would like to generate a random variable in the range [*a*,*b*] with PDF *q*?

Solution:

In this case we get:

$$F' = \pm \frac{1}{q \circ F}$$

Integrating to get $Q'(x)=q(x)$:

$$F' = \pm \frac{1}{Q' \circ F} \Rightarrow \frac{1}{F' \circ F^{-1}} = \pm Q'$$

Sampling

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

$$\left(F^{-1}\right)' = \frac{1}{F' \circ F^{-1}}$$

<u>Challenge</u>:

How about if we would like to generate a random variable in the range [a,b] with PDF q?

<u>Solution</u>:

In this case we get:

$$F' = \pm \frac{1}{q \circ F}$$

Integrating to get $Q'(x)=q(x)$:

$$F' = \pm \frac{1}{Q' \circ F} \Rightarrow \frac{1}{F' \circ F^{-1}} = \pm Q' \Rightarrow \left(F^{-1}\right)' = \pm Q'$$

Sampling

$$q(F(x)) = \frac{p(x)}{|J_F(x)|}$$

$$\left(F^{-1}\right)' = \frac{1}{F' \circ F^{-1}}$$

<u>Challenge</u>:

How about if we would like to generate a random variable in the range [*a*,*b*] with PDF *q*?

<u>Solution</u>:

In this case we get:

$$F' = \pm \frac{1}{q \circ F}$$

Integrating to get $Q'(x)=q(x)$:

$$F' = \pm \frac{1}{Q' \circ F} \Rightarrow \frac{1}{F' \circ F^{-1}} = \pm Q' \Rightarrow \left(F^{-1}\right)' = \pm Q' \Rightarrow F^{-1} = \pm Q + c$$

Sampling

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

$$\left(F^{-1}\right)' = \frac{1}{F' \circ F^{-1}}$$

<u>Challenge</u>:

How about if we would like to generate a random variable in the range [*a*,*b*] with PDF *q*?

<u>Solution</u>:

In this case we get:

$$F' = \pm \frac{1}{q \circ F}$$

Integrating to get $Q'(x)=q(x)$:

$$F' = \pm \frac{1}{Q' \circ F} \Rightarrow \frac{1}{F' \circ F^{-1}} = \pm Q' \Rightarrow \left(F^{-1}\right)' = \pm Q' \Rightarrow F^{-1} = \pm Q + c \Rightarrow F = \pm\left(Q + c\right)^{-1}$$

Sampling

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

$$\left(F^{-1}\right)' = \frac{1}{F' \circ F^{-1}}$$

<u>Challenge</u>:

How about if we would like to generate a random variable in the range [*a*,*b*] with PDF *q*?

<u>Solution</u>:

In this case we get:

$$F' = \pm \frac{1}{q \circ F} \quad \Rightarrow \quad F = \pm\left(Q + c\right)^{-1}$$

$$\Downarrow$$

$$F(\xi) = \begin{Bmatrix} Q^{-1}(\xi) \\ -\left(Q - 1\right)^{-1}(\xi) \end{Bmatrix}$$

$$q\big(F(x)\big) = \frac{p(x)}{\left|J_F(x)\right|}$$

Sampling

$$F(\xi) = \begin{cases} Q^{-1}(\xi) \\ -\left(Q-1\right)^{-1}(\xi) \end{cases}$$

Examples:

If $q(x)=cx^n$ on the domain $[a,b]$

1. We normalize:

$$1 = \int_a^b q(x)dx = c\frac{1}{n+1}\left(b^{n+1} - a^{n+1}\right)$$

$$\Downarrow$$

$$c = \frac{1}{\int_a^b q(x)dx} = \frac{n+1}{b^{n+1} - a^{n+1}}$$

Sampling

$$q\left(F(x)\right)=\frac{p(x)}{\left|J_F(x)\right|}$$

$$F(\xi)=\begin{Bmatrix} Q^{-1}(\xi) \\ -\left(Q-1\right)^{-1}(\xi) \end{Bmatrix}$$

<u>Examples</u>:

If $q(x)=cx^n$ on the domain $[a,b]$

1. We normalize:

$$q(x)=\frac{n+1}{b^{n+1}-a^{n+1}}x^n$$

2. Integrate to get the CDF:

$$Q(x)=\int_a^x \frac{n+1}{b^{n+1}-a^{n+1}}y^n dy=\frac{x^{n+1}-a^{n+1}}{b^{n+1}-a^{n+1}}$$

Sampling

$$q(F(x)) = \frac{p(x)}{|J_F(x)|}$$

$$F(\xi) = \begin{Bmatrix} Q^{-1}(\xi) \\ -(Q-1)^{-1}(\xi) \end{Bmatrix}$$

Examples:

If $q(x)=cx^n$ on the domain $[a,b]$

1. We normalize:

$$q(x) = \frac{n+1}{b^{n+1}-a^{n+1}} x^n$$

2. Integrate to get the CDF:

$$Q(x) = \frac{x^{n+1}-a^{n+1}}{b^{n+1}-a^{n+1}}$$

3. Invert the CDF:

$$F(\xi) = \begin{Bmatrix} \left(\left(b^{n+1}-a^{n+1}\right)\xi + a^{n+1}\right)^{1/n+1} \\ -\left(\left(b^{n+1}-a^{n+1}\right)(\xi-1)+a^{n+1}\right)^{1/n+1} \end{Bmatrix}$$

Sampling

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

$$F(\xi) = \begin{Bmatrix} Q^{-1}(\xi) \\ -\left(Q-1\right)^{-1}(\xi) \end{Bmatrix}$$

<u>Examples</u>:

Uniform random samples in a hemi-sphere:

$$q(\omega) = \frac{1}{2\pi}$$

Sampling

$$q\big(F(x)\big) = \frac{p(x)}{\left|J_F(x)\right|}$$

$$F(\xi) = \begin{cases} Q^{-1}(\xi) \\ -\big(Q-1\big)^{-1}(\xi) \end{cases}$$

Examples:

Uniform random samples in a hemi-sphere:

$$q(\omega) = \frac{1}{2\pi}$$

In terms of the spherical parameterization:

$$\Phi(\theta,\phi) = \big(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta\big)$$

this gives:

$$p(\theta,\phi) = q\big(\Phi(\theta,\phi)\big)\left|J_\Phi(\theta,\phi)\right| = \frac{\sin\theta}{2\pi}$$

$$q\left(F(x)\right)=\frac{p(x)}{\left|J_F(x)\right|}$$

Sampling

$$F(\xi)=\begin{Bmatrix} Q^{-1}(\xi) \\ -\left(Q-1\right)^{-1}(\xi) \end{Bmatrix}$$

Examples:

Uniform random samples in a hemi-sphere:

$$p(\theta,\phi)=\frac{\sin\theta}{2\pi}$$

Recall:

The probability $p(\theta,\phi)$ is the product of the marginal and conditional probabilities:

$$p(\theta,\phi)=p(\theta)\cdot p(\phi\,|\,\theta)$$

Sampling

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

$$F(\xi) = \begin{Bmatrix} Q^{-1}(\xi) \\ -\left(Q-1\right)^{-1}(\xi) \end{Bmatrix}$$

<u>Examples</u>:

Uniform random samples in a hemi-sphere:

$$p(\theta,\phi) = \frac{\sin\theta}{2\pi}$$

The marginal probability of choosing θ is:

$$p(\theta) = \int_0^{2\pi} p(\theta,\phi)d\phi = \int_0^{2\pi} \frac{\sin\theta}{2\pi} d\phi = \sin\theta$$

Sampling

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

$$F(\xi) = \begin{cases} Q^{-1}(\xi) \\ -\left(Q-1\right)^{-1}(\xi) \end{cases}$$

Examples:

Uniform random samples in a hemi-sphere:

$$p(\theta,\phi) = \frac{\sin\theta}{2\pi}$$

The marginal probability of choosing θ is:

$$p(\theta) = \int_0^{2\pi} p(\theta,\phi)d\phi = \int_0^{2\pi} \frac{\sin\theta}{2\pi} d\phi = \sin\theta$$

The conditional probability of choosing ϕ is:

$$p(\phi\,|\,\theta) = \frac{p(\theta,\phi)}{p(\theta)} = \frac{1}{2\pi}$$

Sampling

$$q\big(F(x)\big) = \frac{p(x)}{\left|J_F(x)\right|}$$

$$F(\xi) = \begin{Bmatrix} Q^{-1}(\xi) \\ -\big(Q-1\big)^{-1}(\xi) \end{Bmatrix}$$

Examples:

Uniform random samples in a hemi-sphere:

$$p(\theta,\phi) = \frac{\sin\theta}{2\pi}$$

$$p(\theta) = \sin\theta \qquad p(\phi\,|\,\theta) = \frac{1}{2\pi}$$

Integrating, we get the CDFs:

$$P(\theta) = \int_0^\theta \sin(\theta')d\theta' = 1-\cos\theta \qquad P(\phi\,|\,\theta) = Q(\phi) = \frac{\phi}{2\pi}$$

$$q\big(F(x)\big) = \frac{p(x)}{\left|J_F(x)\right|}$$

Sampling

$$F(\xi) = \begin{Bmatrix} Q^{-1}(\xi) \\ -\big(Q-1\big)^{-1}(\xi) \end{Bmatrix}$$

<u>Examples</u>:

Uniform random samples in a hemi-sphere:

$$p(\theta,\phi) = \frac{\sin\theta}{2\pi}$$

$$P(\theta) = 1 - \cos\theta \qquad\qquad P(\phi\,|\,\theta) = Q(\phi) = \frac{\phi}{2\pi}$$

Inverting, we get:

$$\theta = -\big(P(\xi_1) - 1\big)^{-1} = \cos^{-1}\xi_1 \qquad\qquad \phi = \big(Q(\xi_1) - 1\big)^{-1} = 2\pi\xi_2$$

Sampling

$$q\left(F(x)\right) = \frac{p(x)}{\left|J_F(x)\right|}$$

$$F(\xi) = \begin{Bmatrix} Q^{-1}(\xi) \\ -\left(Q-1\right)^{-1}(\xi) \end{Bmatrix}$$

<u>Examples</u>:

Uniform random samples in a hemi-sphere:

$$\theta = \cos^{-1} \xi_1 \qquad \phi = 2\pi\xi_2$$

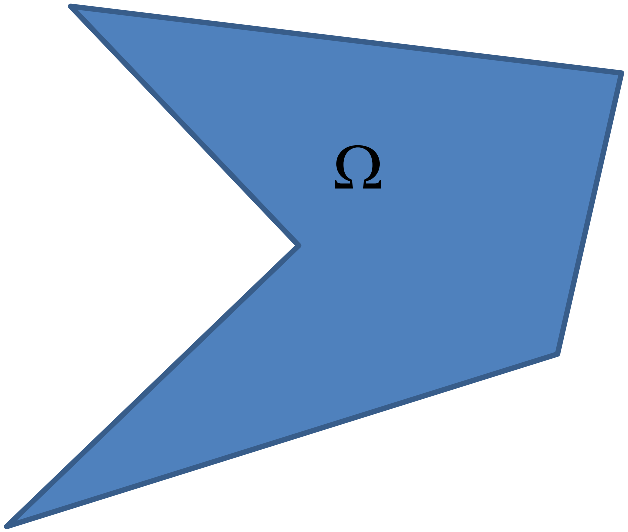
<u>Note</u>:

Even though this transforms uniform random variables in $[0,1]^2$ to uniform random variables on H^2, the distribution does not preserve areas, so stratification in $[0,1]^2$ does not guarantee stratification on H^2.

Rejection Sampling

Challenge:

What happens when Ω is complicated?



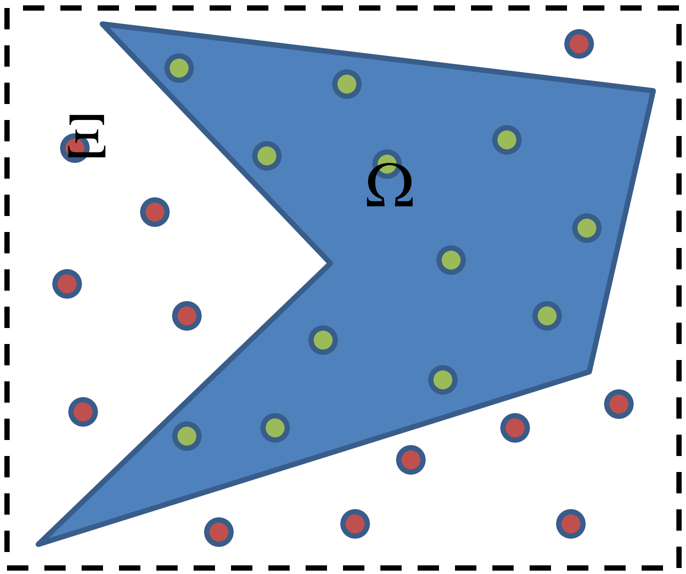
Rejection Sampling

Challenge:

What happens when Ω is complicated?

Solution:

Generate samples in a simpler domain $\Xi \supset \Omega$ and discard samples that are not in Ω.



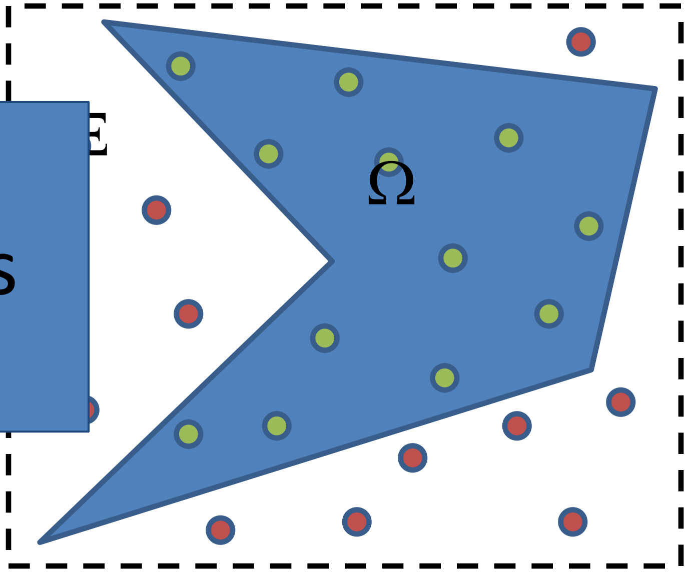
Rejection Sampling

Challenge:

What happens when Ω is complicated?

Solution:

Generate samples in a simpler domain $\Xi \supset \Omega$ and discard samples that are not in Ω.



Note:

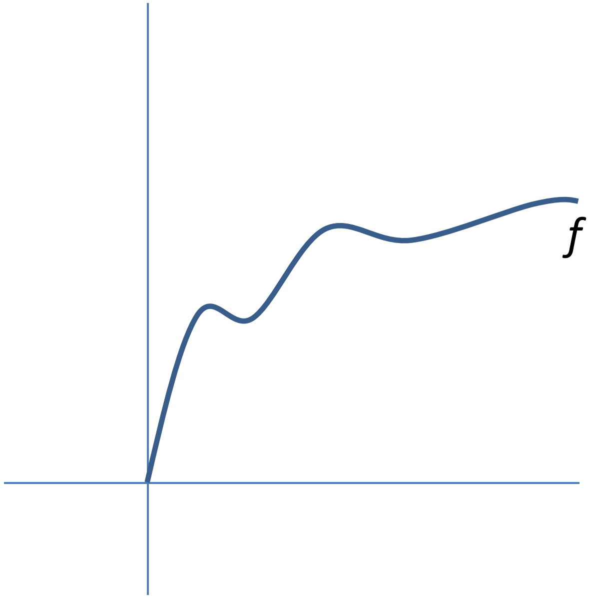
Effectiveness of the sampling is

$$\text{Area}(\Omega)/\text{Area}(\Xi)$$

Rejection Sampling

Challenge:

How do we sample according to a function *f* that is complicated (e.g. unknown PDF)?



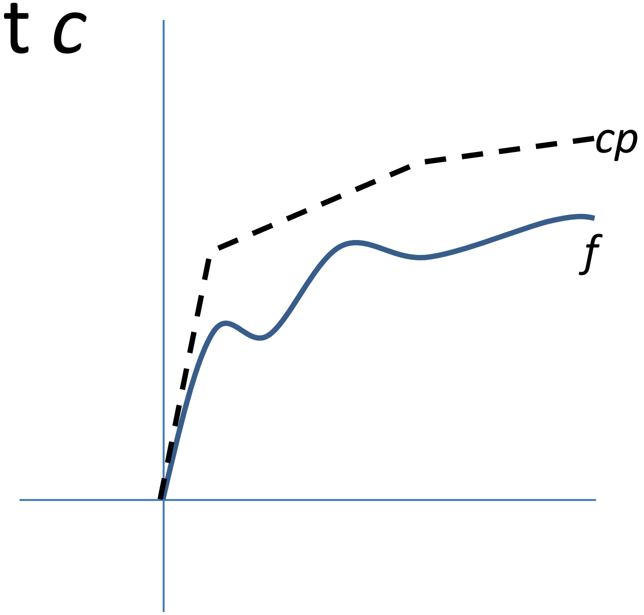
Rejection Sampling

<u>Challenge</u>:

How do we sample according to a function f that is complicated (e.g. unknown PDF)?

<u>Solution</u>:

Find a simple PDF p and constant c such that $f \leq cp$:



Rejection Sampling

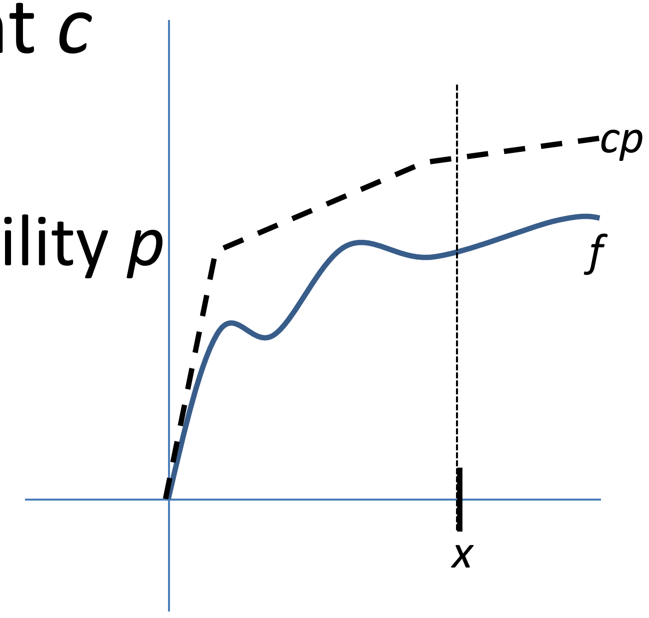
<u>Challenge</u>:

How do we sample according to a function *f* that is complicated (e.g. unknown PDF)?

<u>Solution</u>:

Find a simple PDF *p* and constant *c* such that $f \leq cp$:

- Generate sample *x* with probability *p*



Rejection Sampling

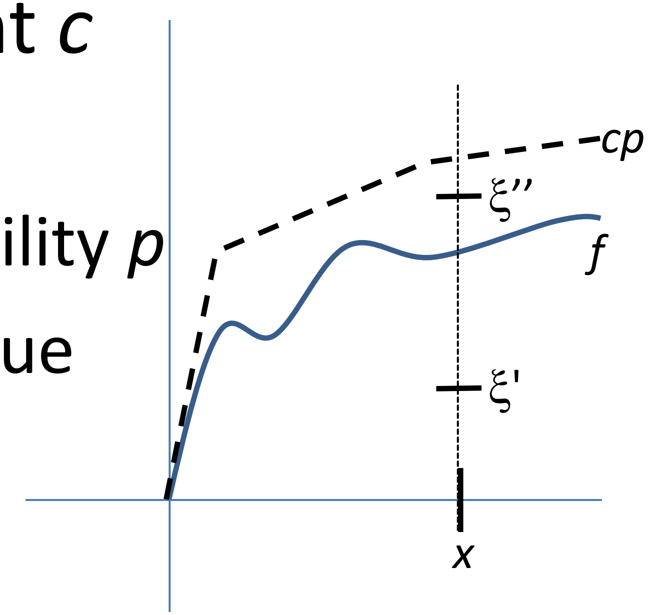
<u>Challenge</u>:

How do we sample according to a function f that is complicated (e.g. unknown PDF)?

<u>Solution</u>:

Find a simple PDF p and constant c such that $f \leq cp$:

- Generate sample x with probability p
- Generate a uniform random value $\xi \in [0, cp(x)]$.



Rejection Sampling

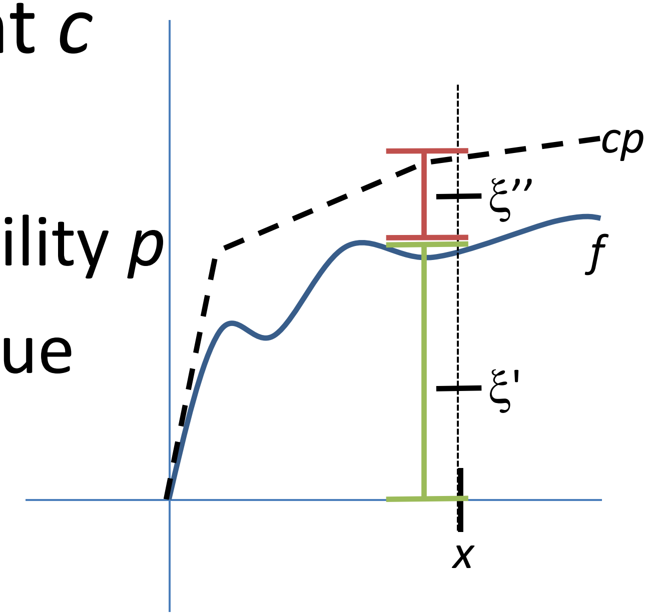
Challenge:

How do we sample according to a function f that is complicated (e.g. unknown PDF)?

Solution:

Find a simple PDF p and constant c such that $f \leq cp$:

- Generate sample x with probability p
- Generate a uniform random value $\xi \in [0, cp(x)]$
- Keep x if $\xi < f(x)$



Rejection Sampling

<u>Challenge</u>:

How do we sample according to a function *f* that is complicated (e.g. unknown PDF)?

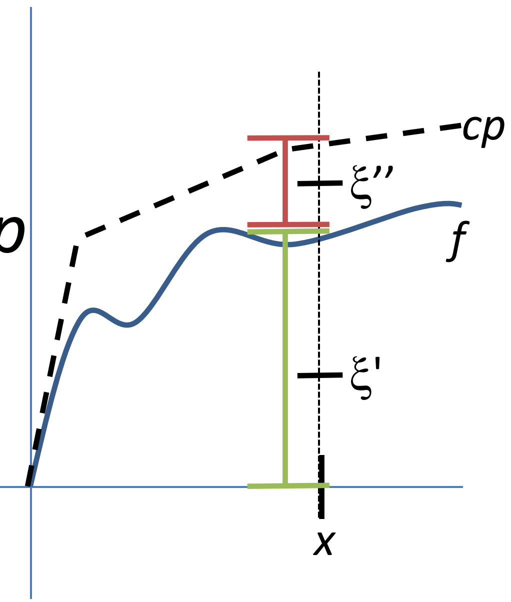
<u>Solution</u>:

Find a simple PDF *p* and constant *c* such that $f \leq cp$:

<u>Note</u>:

Effectiveness of the sampling is

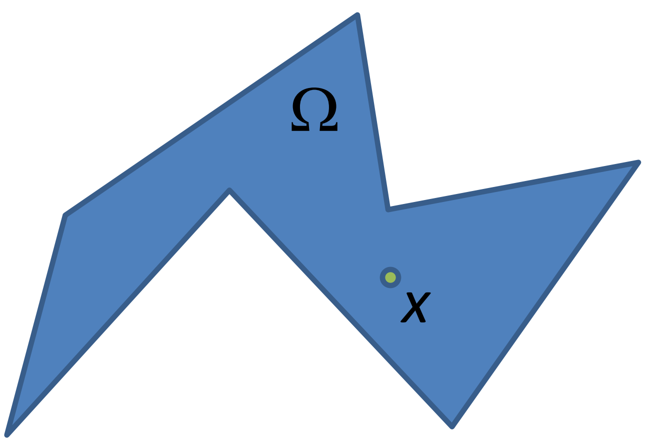
$$\int_\Omega f(x)p(x)dx \Big/ \int_\Omega cp(x)p(x)dx$$



Metropolis Sampling

To generate samples with probability proportional to a function $f \geq 0$:

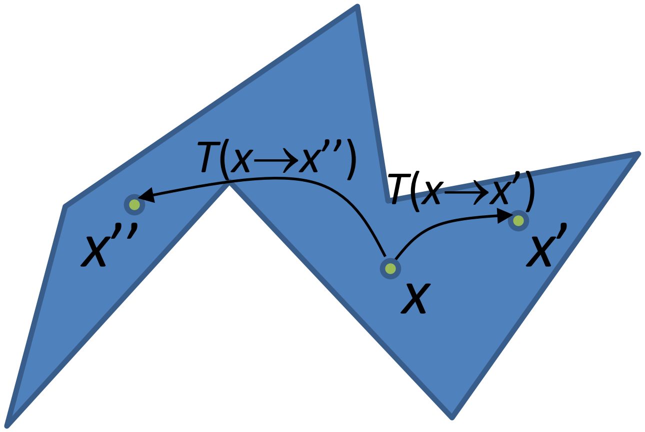
1. Start with a random sample $x \in \Omega$.



Metropolis Sampling

To generate samples with probability proportional to a function $f \geq 0$:

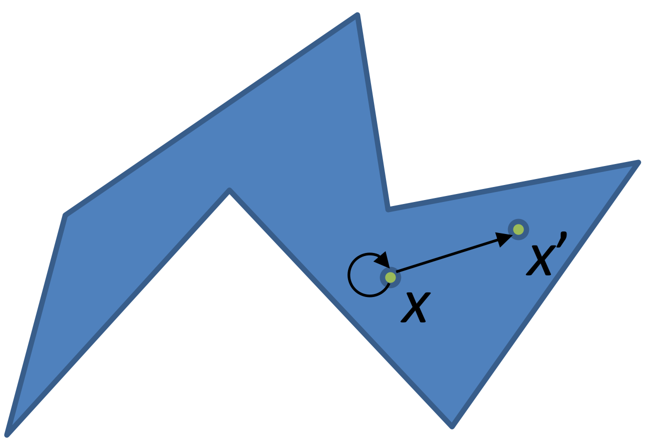
1. Start with a random sample $x \in \Omega$.
2. Generate a random mutation x' under the PDF $T(x \to x')$.



Metropolis Sampling

To generate samples with probability proportional to a function $f \geq 0$:

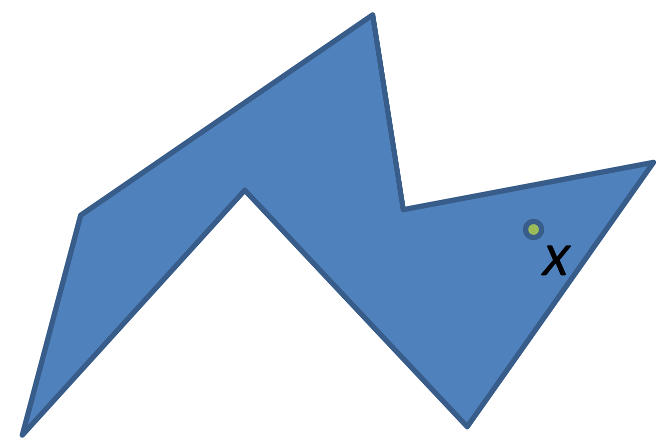
1. Start with a random sample $x \in \Omega$.
2. Generate a random mutation x' under the PDF $T(x \to x')$.
3. With probability a, set $x = x'$.



Metropolis Sampling

To generate samples with probability proportional to a function $f \geq 0$:

1. Start with a random sample $x \in \Omega$.
2. Generate a random mutation x' under the PDF $T(x \rightarrow x')$.
3. With probability a, set $x = x'$.
4. Go to step 2.



Metropolis Sampling

Challenge 1:

Given the transition probability T, how should we define the acceptance probability a so that the distribution is proportional to f?

Metropolis Sampling

<u>Solution</u>:

Consider the case when there are only two states, $\Omega=\{x_1,x_2\}$.

Metropolis Sampling

Solution:

Consider the case when there are only two states, $\Omega=\{x_1,x_2\}$.

If the acceptance probability is set up correctly, the probability of transitioning to x_1 should be equal to the probability of choosing x_1:

$$p_1 = \underbrace{p_1 \cdot \left[T_{12} \cdot (1-a_{12}) + T_{11}\right]}_{\text{Probability of starting with } x_1 \text{ and not mutating to } x_2.} + \underbrace{p_2 \cdot T_{21} \cdot a_{21}}_{\text{Probability of starting with } x_2 \text{ and mutating to } x_1.}$$

with:

$$f_i = f(x_i),\ \ T_{ij} = T(x_i \rightarrow x_j),\ \ a_{ij} = a(x_i \rightarrow x_j),\ \ \alpha = \frac{1}{f_1 + f_2},\ \ p_i = p(x_i) = \alpha f_i$$

Metropolis Sampling

<u>Solution</u>:

Consider the case when there are only two states, $\Omega=\{x_1,x_2\}$.

If the acceptance probability is set up correctly, the probability of transitioning to x_1 should be equal to the probability of choosing x_1:

$$p_1 = p_1 \cdot \left[T_{12} \cdot (1 - a_{12}) + T_{11}\right] + \ p_2 \cdot T_{21} \cdot a_{21}$$

$$f_1 = f_1 \cdot T_{12} - f_1 \cdot T_{12} \cdot a_{12} + f_1 \cdot T_{11} \ + \ f_2 \cdot T_{21} \cdot a_{21}$$

Metropolis Sampling

<u>Solution</u>:

Consider the case when there are only two states, $\Omega=\{x_1,x_2\}$.

If the acceptance probability is set up correctly, the probability of transitioning to x_1 should be equal to the probability of choosing x_1:

$$p_1 = p_1 \cdot \left[T_{12} \cdot (1 - a_{12}) + T_{11}\right] + \; p_2 \cdot T_{21} \cdot a_{21}$$

$$f_1 = f_1 \cdot T_{12} - f_1 \cdot T_{12} \cdot a_{12} + f_1 \cdot T_{11} \; + \; f_2 \cdot T_{21} \cdot a_{21}$$

$$f_1 = f_1 \cdot T_{12} - f_1 \cdot T_{12} \cdot a_{12} + f_1 \cdot (1 - T_{12}) \; + \; f_2 \cdot T_{21} \cdot a_{21}$$

Metropolis Sampling

<u>Solution</u>:

Consider the case when there are only two states, $\Omega=\{x_1,x_2\}$.

If the acceptance probability is set up correctly, the probability of transitioning to x_1 should be equal to the probability of choosing x_1:

$$p_1 = p_1 \cdot \left[T_{12} \cdot (1 - a_{12}) + T_{11}\right] + \ p_2 \cdot T_{21} \cdot a_{21}$$

$$f_1 = f_1 \cdot T_{12} - f_1 \cdot T_{12} \cdot a_{12} + f_1 \cdot T_{11} \ + \ f_2 \cdot T_{21} \cdot a_{21}$$

$$f_1 = f_1 \cdot T_{12} - f_1 \cdot T_{12} \cdot a_{12} + f_1 \cdot (1 - T_{12}) \ + \ f_2 \cdot T_{21} \cdot a_{21}$$

$$0 = - f_1 \cdot T_{12} \cdot a_{12} \ + \ f_2 \cdot T_{21} \cdot a_{21}$$

Metropolis Sampling

Solution:

Consider the case when there are only two states, $\Omega=\{x_1,x_2\}$.

Thus, the acceptance probability should satisfy:

$$f_1 \cdot T_{12} \cdot a_{12} = f_2 \cdot T_{21} \cdot a_{21}$$

Metropolis Sampling

Solution:

Consider the case when there are only two states, $\Omega=\{x_1,x_2\}$.

Thus, the acceptance probability should satisfy:

$$f_1 \cdot T_{12} \cdot a_{12} = f_2 \cdot T_{21} \cdot a_{21}$$

Note that this ensures stationarity of p but does not ensure convergence or uniqueness:

1. Uniqueness: Setting $a_{12}=a_{21}=0$ can converge to the wrong solution (reducible).
2. Convergence: If $T_{11}=T_{22}=0$ and $a_{12}=a_{21}=1$, may not converge (cyclic).

Metropolis Sampling

Solution:

Consider the case when there are only two states, $\Omega=\{x_1,x_2\}$.

Thus, the acceptance probability should satisfy:

$$f_1 \cdot T_{12} \cdot a_{12} = f_2 \cdot T_{21} \cdot a_{21}$$

Assuming T_{ij}>0, we can get uniqueness and convergence by setting:

$$a(x_1 \to x_2) = \min\left(1, \frac{f(x_2)\cdot T(x_2 \to x_1)}{f(x_1)\cdot T(x_1 \to x_2)}\right)$$

Metropolis Sampling

<u>Solution</u>:

More generally, if the state space is discrete, the probability of choosing state x_j is:

$$p_j = \underbrace{p_j \cdot T_{jj} + p_j \cdot \left[\sum_{i \neq j} T_{ji} \cdot (1 - a_{ji}) \right]}_{\text{Probability of starting with } x_j \text{ and not mutating out.}} + \underbrace{\sum_{i \neq j} p_j \cdot T_{ij} \cdot a_{ij}}_{\text{Probability of not starting as } x_j \text{ and mutating to } x_j.}$$

Metropolis Sampling

Solution:

More generally, if the state space is discrete, the probability of choosing state x_j is:

$$p_j = p_j \cdot T_{jj} + p_j \cdot \left[\sum_{i \neq j} T_{ji} \cdot (1 - a_{ji}) \right] + \sum_{i \neq j} p_j \cdot T_{ij} \cdot a_{ij}$$

$$f_j = f_j \cdot \left(1 - \sum_{i \neq j} T_{ji} \right) + \sum_{i \neq j} f_j \cdot T_{ji} \cdot (1 - a_{ji}) + \sum_{i \neq j} f_i \cdot T_{ij} \cdot a_{i1}$$

Metropolis Sampling

<u>Solution</u>:

More generally, if the state space is discrete, the probability of choosing state x_j is:

$$p_j = p_j \cdot T_{jj} + p_j \cdot \left[\sum_{i \neq j} T_{ji} \cdot (1 - a_{ji}) \right] + \sum_{i \neq j} p_j \cdot T_{ij} \cdot a_{ij}$$

$$f_j = f_j \cdot \left(1 - \sum_{i \neq j} T_{ji} \right) + \sum_{i \neq j} f_j \cdot T_{ji} \cdot (1 - a_{ji}) + \sum_{i \neq j} f_i \cdot T_{ij} \cdot a_{ij}$$

$$f_j = f_j - \sum_{i \neq j} f_j \cdot T_{ji} + \sum_{i \neq j} f_j \cdot T_{ji} - \sum_{i \neq j} f_j \cdot T_{ji} \cdot a_{ji} + \sum_{i \neq j} f_i \cdot T_{ij} \cdot a_{ij}$$

Metropolis Sampling

Solution:

More generally, if the state space is discrete, the probability of choosing state x_j is:

$$p_j = p_j \cdot T_{jj} + p_j \cdot \left[\sum_{i \neq j} T_{ji} \cdot (1 - a_{ji}) \right] + \sum_{i \neq j} p_j \cdot T_{ij} \cdot a_{ij}$$

$$f_j = f_j \cdot \left(1 - \sum_{i \neq j} T_{ji} \right) + \sum_{i \neq j} f_j \cdot T_{ji} \cdot (1 - a_{ji}) + \sum_{i \neq j} f_i \cdot T_{ij} \cdot a_{ij}$$

$$f_j = f_j - \sum_{i \neq j} f_j \cdot T_{ji} + \sum_{i \neq j} f_j \cdot T_{ji} - \sum_{i \neq j} f_j \cdot T_{ji} \cdot a_{ji} + \sum_{i \neq j} f_i \cdot T_{ij} \cdot a_{ij}$$

$$\sum_{i \neq j} f_j \cdot T_{ji} \cdot a_{ji} = \sum_{i \neq j} f_i \cdot T_{ij} \cdot a_{ij}$$

Metropolis Sampling

<u>Solution</u>:

More generally, if the state space is discrete, the probability of choosing state x_j is:

$$\sum_{i \neq j} f_j \cdot T_{ji} \cdot a_{ji} = \sum_{i \neq j} f_i \cdot T_{ij} \cdot a_{ij}$$

which is satisfied if we ensure that:

$$f_j \cdot T_{ji} \cdot a_{ji} = f_i \cdot T_{ij} \cdot a_{ij} \quad \forall i \neq j$$

Metropolis Sampling

<u>Solution</u>:

More generally, if the state space is discrete, the probability of choosing state x_j is:

$$\sum_{i \neq j} f_j \cdot T_{ji} \cdot a_{ji} = \sum_{i \neq j} f_i \cdot T_{ij} \cdot a_{ij}$$

which is satisfied if we ensure that:

$$f_j \cdot T_{ji} \cdot a_{ji} = f_i \cdot T_{ij} \cdot a_{ij} \quad \forall i \neq j$$

This not only ensures that mutating *p* we get back *p*, but also that repeatedly mutating and PDF *q* we get a multiple of *p*.

Metropolis Sampling

Challenge 2:

How should we choose the initial state *x*?

Metropolis Sampling

Challenge 2:

How should we choose the initial state *x*?

Solution (Burn In):

Run for many iterations and discard the initial samples.

Metropolis Sampling

<u>Solution</u>:

More generally, we can think of Metropolis sampling as taking some PDF *p* and mutating it into a new PDF *q*:

$$q_i = \sum_j p_j \cdot T_{ji} \cdot a_{ji}$$

Metropolis Sampling

<u>Solution</u>:

More generally, we can think of Metropolis sampling as taking some PDF p and mutating it into a new PDF q:

$$q_i = \sum_j p_j \cdot T_{ji} \cdot a_{ji}$$

Setting M be the matrix $M_{ij}=T_{ji}a_{ji}$, and setting $p=\{p_1,\ldots,p_n\}$ and $q=\{q_1,\ldots,q_n\}$ this becomes:

$$q = Mp$$

Metropolis Sampling

$$q = Mp$$

Conditions on *M*:

1. Stationarity: Applying *M* to *f* we should get back *f*: *Mf*=*f*. (*f* is an eigenvector of *M* with eigenvalue 1).
2. Convergence: Repeatedly applying *M* to a PDF *p* we should converge to something. (All eigenvalues of *M* have magnitude $\leq$1.)
3. Uniqueness: That "something" should be a scalar multiple of *f*. (All other eigenvalues of *M* have magnitude <1.)