Statistics for Astronomers: Lecture 19, 2021.01.14

Prof. Sundar Srinivasan

IRyA/UNAM

Review

Prior and posterior predictive distributions.
Jeffreys prior.

Jupyter demos

1. ▸ Download this Jupyter notebook.
2. ▸ Navigate to Colaboratory.
3. Sign in
4. Click on "Upload" and upload the notebook you downloaded in step 1.

Sampling techniques

Inverse-transform sampling ("CDF method")

If $X \sim p_X(x)$ and $y = F_X(x)$, then $y \sim \mathrm{Uniform}$.

To draw samples from $p_X(x)$, we can draw y from a Uniform distribution and then compute x.

Example: draw samples from an exponential distribution, $X \sim e^{-x}$ for $0 \leq x < \infty$.

$$\text{CDF: } F_X(x) = \int_0^x dx' \; e^{-x'} = 1 - e^{-x}.$$

$$\text{Set } y = F_X(x) = 1 - e^{-x} \implies x = -\ln(1-y).$$

$$x \in [0, \infty) \implies y \in [0, 1].$$

For $x \sim e^{-x}$, draw $y \sim \mathrm{Uniform}(0,1)$ and set $x = -\ln(1-y)$.

Can't use if $F_X(x)$ difficult or impossible to invert.

Rejection sampling

Aim: to sample from a target distribution $p(x)$ which isn't straightforward.

Rejection sampling

Aim: to sample from a target distribution $p(x)$ which isn't straightforward.
Workaround: sample from a proposal distribution $g(x)$ that envelopes $p(x)$ ($g(x) \geq p(x)$ over relevant range).
If $g(x)$ normalised, can't be $\geq p(x)$ over entire range for x.
Set $g(x) = M\ h(x)$ where $h(x)$ is normalised and $M > 1$.

Rejection sampling

Aim: to sample from a target distribution $p(x)$ which isn't straightforward.
Workaround: sample from a proposal distribution $g(x)$ that envelopes $p(x)$
$(g(x) \geq p(x)$ over relevant range).
If $g(x)$ normalised, can't be $\geq p(x)$ over entire range for x.
Set $g(x) = M\ h(x)$ where $h(x)$ is normalised and $M > 1$.

Rejection sampling procedure:

Rejection sampling

Aim: to sample from a target distribution $p(x)$ which isn't straightforward.
Workaround: sample from a proposal distribution $g(x)$ that envelopes $p(x)$ ($g(x) \geq p(x)$ over relevant range).
If $g(x)$ normalised, can't be $\geq p(x)$ over entire range for x.
Set $g(x) = M\ h(x)$ where $h(x)$ is normalised and $M > 1$.

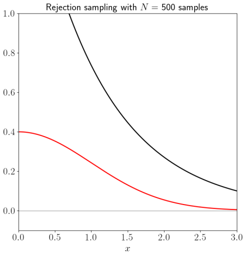
Ex.: Draw samples from $p(x)$ where

$$p(x) = \sqrt{\frac{2}{\pi}}\ e^{-x^2/2}, 0 \leq x < \infty.$$

$$h(x) = e^{-x}, 0 \leq x < \infty; \quad g(x) = 2\ h(x).$$

Rejection sampling procedure:

1. Find $h(x)$ and some $M > 1$ such that $M\ h(x) \equiv g(x) \geq p(x)$.



Rejection sampling

Aim: to sample from a target distribution $p(x)$ which isn't straightforward.
Workaround: sample from a proposal distribution $g(x)$ that envelopes $p(x)$ ($g(x) \geq p(x)$ over relevant range).
If $g(x)$ normalised, can't be $\geq p(x)$ over entire range for x.
Set $g(x) = M\ h(x)$ where $h(x)$ is normalised and $M > 1$.

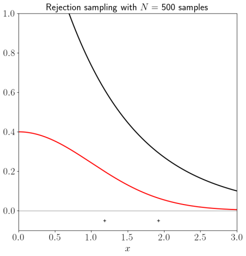
Ex.: Draw samples from $p(x)$ where

$$p(x) = \sqrt{\frac{2}{\pi}}\, e^{-x^2/2}, 0 \leq x < \infty.$$

$h(x) = e^{-x}, 0 \leq x < \infty$; $\quad g(x) = 2\ h(x)$.
Draws from $h(x)$: inverse-transform sampling.

Rejection sampling procedure:

1. Find $h(x)$ and some $M > 1$ such that $M\ h(x) \equiv g(x) \geq p(x)$.
2. Draw samples $\{x_i\}$ from the proposal distribution $g(x)$.



Rejection sampling

Aim: to sample from a target distribution $p(x)$ which isn't straightforward.
Workaround: sample from a proposal distribution $g(x)$ that envelopes $p(x)$ ($g(x) \geq p(x)$ over relevant range).
If $g(x)$ normalised, can't be $\geq p(x)$ over entire range for x.
Set $g(x) = M\ h(x)$ where $h(x)$ is normalised and $M > 1$.

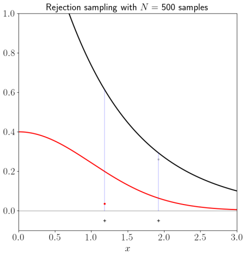
Ex.: Draw samples from $p(x)$ where

$$p(x) = \sqrt{\frac{2}{\pi}}\ e^{-x^2/2}, 0 \leq x < \infty.$$

$h(x) = e^{-x}, 0 \leq x < \infty;\quad g(x) = 2\ h(x)$.
Draws from $h(x)$: inverse-transform sampling.

Rejection sampling procedure:

1. Find $h(x)$ and some $M > 1$ such that $M\ h(x) \equiv g(x) \geq p(x)$.
2. Draw samples $\{x_i\}$ from the proposal distribution $g(x)$.
3. For each x_i, draw y_i from $\mathrm{Uniform}[0, g(x_i)]$ (find a value between 0 and the value of the proposal distribution at x_i).



Rejection sampling

Aim: to sample from a target distribution $p(x)$ which isn't straightforward.
Workaround: sample from a proposal distribution $g(x)$ that envelopes $p(x)$ ($g(x) \geq p(x)$ over relevant range).
If $g(x)$ normalised, can't be $\geq p(x)$ over entire range for x.
Set $g(x) = M\ h(x)$ where $h(x)$ is normalised and $M > 1$.

Ex.: Draw samples from $p(x)$ where

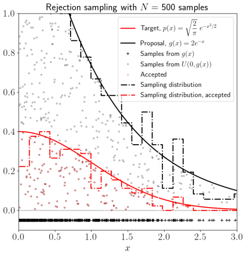
$$p(x) = \sqrt{\frac{2}{\pi}}\ e^{-x^2/2}, 0 \leq x < \infty.$$

$h(x) = e^{-x}, 0 \leq x < \infty;\quad g(x) = 2\ h(x)$.

Draws from $h(x)$: inverse-transform sampling.

Rejection sampling procedure:

1. Find $h(x)$ and some $M > 1$ such that $M\ h(x) \equiv g(x) \geq p(x)$.
2. Draw samples $\{x_i\}$ from the proposal distribution $g(x)$.
3. For each x_i, draw y_i from $\mathrm{Uniform}[0, g(x_i)]$ (find a value between 0 and the value of the proposal distribution at x_i).
4. Reject all pairs (x_i, y_i) such that $y_i \geq p(x_i)$.



Rejection sampling

Aim: to sample from a target distribution $p(x)$ which isn't straightforward.
Workaround: sample from a proposal distribution $g(x)$ that envelopes $p(x)$ ($g(x) \geq p(x)$ over relevant range).
If $g(x)$ normalised, can't be $\geq p(x)$ over entire range for x.
Set $g(x) = M\ h(x)$ where $h(x)$ is normalised and $M > 1$.

Ex.: Draw samples from $p(x)$ where

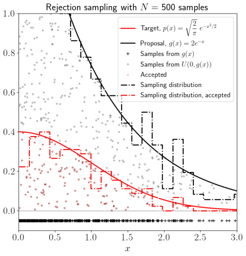
$$p(x) = \sqrt{\frac{2}{\pi}}\ e^{-x^2/2}, 0 \leq x < \infty.$$

$h(x) = e^{-x}, 0 \leq x < \infty;\quad g(x) = 2\ h(x).$

Draws from $h(x)$: inverse-transform sampling.

Rejection sampling procedure:

1. Find $h(x)$ and some $M > 1$ such that $M\ h(x) \equiv g(x) \geq p(x)$.
2. Draw samples $\{x_i\}$ from the proposal distribution $g(x)$.
3. For each x_i, draw y_i from $\mathrm{Uniform}[0, g(x_i)]$ (find a value between 0 and the value of the proposal distribution at x_i).
4. Reject all pairs (x_i, y_i) such that $y_i \geq p(x_i)$.



$$P(\mathrm{accept}) = \frac{\int dx\ p(x)}{\int dx\ g(x)} = \frac{\int dx\ p(x)}{\int dx\ M\ h(x)} = \frac{1}{M}.$$

Most efficient when smallest $M > 1$ chosen!

Monte Carlo

Algorithms to draw random samples from some target distribution.

The sampling may then be applied to a wide variety of problems – *e.g.*, error propagation, integration (quadrature), sampling posterior PDFs from Bayesian analysis for point and interval estimates of parameters.

Usually simple to implement, but inefficient compared to other methods.

Becomes more efficient as number of parameters increases (dimensionality).

For some distributions/problems, the only choice that works.

Application 1: Error propagation

If $f(x)$ is a function of a random variable x and $\mathbb{E}[x] = x_0 \Rightarrow \mathbb{E}[f(x)] = f(x_0)$

Application 1: Error propagation

If $f(x)$ is a function of a random variable x and $\mathbb{E}[x] = x_0 \Rightarrow \mathbb{E}[f(x)] = f(x_0)$

Taylor expansion around $\mathbb{E}[x]$: $f(x) = f(x_0) + \left(\frac{\partial f(x)}{\partial x}\right)_{x_0} (x - x_0) + \left(\frac{\partial^2 f(x)}{\partial x^2}\right)_{x_0} \frac{(x - x_0)^2}{2} + \cdots$

Application 1: Error propagation

If $f(x)$ is a function of a random variable x and $\mathbb{E}[x] = x_0 \Rightarrow \mathbb{E}[f(x)] = f(x_0)$

Taylor expansion around $\mathbb{E}[x]$: $f(x) = f(x_0) + \left(\frac{\partial f(x)}{\partial x}\right)_{x_0} (x - x_0) + \left(\frac{\partial^2 f(x)}{\partial x^2}\right)_{x_0} \frac{(x - x_0)^2}{2} + \cdots$

$\mathrm{Var}[f(x)] \approx \left(\frac{\partial f(x)}{\partial x}\right)_{x_0}^2 \mathrm{Var}[x] \Longrightarrow \sigma_f \approx \left(\frac{\partial f(x)}{\partial x}\right)_{x_0} \sigma_x$ Fails for large rel. unc.

Application 1: Error propagation

If $f(x)$ is a function of a random variable x and $\mathbb{E}[x] = x_0 \Rightarrow \mathbb{E}[f(x)] = f(x_0)$

Taylor expansion around $\mathbb{E}[x]$: $f(x) = f(x_0) + \left(\frac{\partial f(x)}{\partial x}\right)_{x_0} (x - x_0) + \left(\frac{\partial^2 f(x)}{\partial x^2}\right)_{x_0} \frac{(x - x_0)^2}{2} + \cdots$

$\mathrm{Var}[f(x)] \approx \left(\frac{\partial f(x)}{\partial x}\right)^2_{x_0} \mathrm{Var}[x] \Longrightarrow \sigma_f \approx \left(\frac{\partial f(x)}{\partial x}\right)_{x_0} \sigma_x$ Fails for large rel. unc.

Example: $m_{\mathrm{IRAC}\ 8\mu m} = 1.27(1 \pm s/100)$ mag. What is the relative uncertainty in the flux?

$F \propto 10^{-\frac{m}{2.5}} \Longrightarrow$ rel. unc. $\approx 0.012\ s$ (Assumes symm. about $\mathbb{E}[F]$).

$\approx 17\%$ (s=15)

$\approx 52\%$ (s=45)

Application 1: Error propagation

If $f(x)$ is a function of a random variable x and $\mathbb{E}[x] = x_0 \Rightarrow \mathbb{E}[f(x)] = f(x_0)$

Taylor expansion around $\mathbb{E}[x]$: $f(x) = f(x_0) + \left(\frac{\partial f(x)}{\partial x}\right)_{x_0} (x - x_0) + \left(\frac{\partial^2 f(x)}{\partial x^2}\right)_{x_0} \frac{(x - x_0)^2}{2} + \cdots$

$\mathrm{Var}[f(x)] \approx \left(\frac{\partial f(x)}{\partial x}\right)^2_{x_0} \mathrm{Var}[x] \Longrightarrow \sigma_f \approx \left(\frac{\partial f(x)}{\partial x}\right)_{x_0} \sigma_x$ Fails for large rel. unc.

Example: $m_{\mathrm{IRAC}\ 8\mu m} = 1.27(1 \pm s/100)$ mag. What is the relative uncertainty in the flux?

$F \propto 10^{-\frac{m}{2.5}} \Longrightarrow$ rel. unc. $\approx 0.012\ s$ (Assumes symm. about $\mathbb{E}[F]$).

$\approx 17\%$ (s=15)

$\approx 52\%$ (s=45)

Monte Carlo sampling:

Assume m normally distributed about 1.27 mag.

Sample from magnitude distribution, compute flux.

Application 1: Error propagation

If $f(x)$ is a function of a random variable x and $\mathbb{E}[x] = x_0 \Rightarrow \mathbb{E}[f(x)] = f(x_0)$

Taylor expansion around $\mathbb{E}[x]$: $f(x) = f(x_0) + \left(\frac{\partial f(x)}{\partial x}\right)_{x_0} (x - x_0) + \left(\frac{\partial^2 f(x)}{\partial x^2}\right)_{x_0} \frac{(x - x_0)^2}{2} + \cdots$

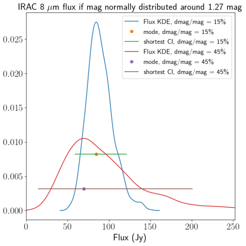
$\mathrm{Var}[f(x)] \approx \left(\frac{\partial f(x)}{\partial x}\right)_{x_0}^2 \mathrm{Var}[x] \Longrightarrow \sigma_f \approx \left(\frac{\partial f(x)}{\partial x}\right)_{x_0} \sigma_x$ Fails for large rel. unc.

Example: $m_{\mathrm{IRAC}\ 8\mu m} = 1.27(1 \pm s/100)$ mag. What is the relative uncertainty in the flux?

$F \propto 10^{-\frac{m}{2.5}} \Longrightarrow$ rel. unc. $\approx 0.012\ s$ (Assumes symm. about $\mathbb{E}[F]$).

$\approx 17\%$ (s=15)

$\approx 52\%$ (s=45)



Monte Carlo sampling:

- Assume m normally distributed about 1.27 mag.
- Sample from magnitude distribution, compute flux.
- For small s, flux distribution normal.
- As $s \uparrow$, distribution becomes skewed.

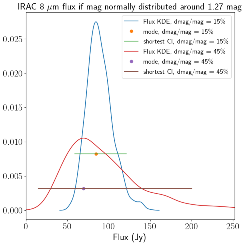
Application 1: Error propagation

If $f(x)$ is a function of a random variable x and $\mathbb{E}[x] = x_0 \Rightarrow \mathbb{E}[f(x)] = f(x_0)$

Taylor expansion around $\mathbb{E}[x]$: $f(x) = f(x_0) + \left(\frac{\partial f(x)}{\partial x}\right)_{x_0} (x - x_0) + \left(\frac{\partial^2 f(x)}{\partial x^2}\right)_{x_0} \frac{(x - x_0)^2}{2} + \cdots$

$\mathrm{Var}[f(x)] \approx \left(\frac{\partial f(x)}{\partial x}\right)_{x_0}^2 \mathrm{Var}[x] \Longrightarrow \sigma_f \approx \left(\frac{\partial f(x)}{\partial x}\right)_{x_0} \sigma_x$ Fails for large rel. unc.

Example: $m_{\mathrm{IRAC}\ 8\mu m} = 1.27(1 \pm s/100)$ mag. What is the relative uncertainty in the flux?



$F \propto 10^{-\frac{m}{2.5}} \Longrightarrow$ rel. unc. $\approx 0.012\, s$ (Assumes symm. about $\mathbb{E}[F]$).

$\approx 17\%$ (s=15)

$\approx 52\%$ (s=45)

Monte Carlo sampling:

Assume m normally distributed about 1.27 mag.

Sample from magnitude distribution, compute flux.

For small s, flux distribution normal.

As $s \uparrow$, distribution becomes skewed.

Useful for (a) simultaneous propagation of many errors and/or (b) nonlinear relationships (*e.g.*, blackbody flux in terms of T).

Application 2: Quadrature for complicated functions

Example: compute the area inside the curve $(x^2 + y^2)^2 + 4x(x^2 + y^2 - \pi y) = y^2$.

Difficult to compute analytically.

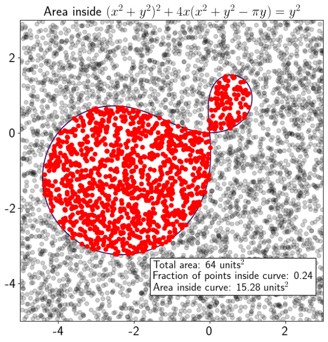
Monte Carlo way: draw random sample of points over a larger shape of known area, then compute fraction of points inside desired area.

Application 2: Quadrature for complicated functions

Example: compute the area inside the curve $(x^2+y^2)^2 + 4x(x^2+y^2-\pi y) = y^2$.

Difficult to compute analytically.

Monte Carlo way: draw random sample of points over a larger shape of known area, then compute fraction of points inside desired area.



Larger area: square of area 64 units^2.

Fraction of points inside desired area $\approx$ 24%.

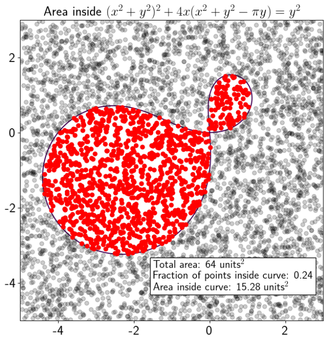
Desired area $\approx$ 15.28 units^2.

Application 2: Quadrature for complicated functions

Example: compute the area inside the curve $(x^2 + y^2)^2 + 4x(x^2 + y^2 - \pi y) = y^2$.

Difficult to compute analytically.

Monte Carlo way: draw random sample of points over a larger shape of known area, then compute fraction of points inside desired area.



Larger area: square of area 64 units^2.

Fraction of points inside desired area $\approx$ 24%.

Desired area $\approx$ 15.28 units^2.

Rejection sampling.

1) Select an area that envelopes desired area,

2) Sample uniformly over the enveloping area,

3) Reject samples generated outside desired area.

Quadrature as an expectation value (simple Monte Carlo)

Evaluate $\int_a^b dx\ g(x)$ using Monte Carlo sampling.

Quadrature as an expectation value (simple Monte Carlo)

Evaluate $\int_a^b dx\ g(x)$ using Monte Carlo sampling.

$$\int_a^b dx\ g(x) = (b-a)\int_a^b dx\ \frac{1}{b-a}\ g(x) = \int_a^b dx\ p_X(x)\ g(x), \text{ with } p_X(x) = \mathrm{Uniform}(a, b).$$

$$= \mathbb{E}[g(x)] \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i), \text{ with the } x_i \text{ drawn from } \mathrm{Uniform}(a, b).$$

Quadrature as an expectation value (simple Monte Carlo)

Evaluate $\int_a^b dx\ g(x)$ using Monte Carlo sampling.

$$\int_a^b dx\ g(x) = (b-a)\int_a^b dx\ \frac{1}{b-a}\ g(x) = \int_a^b dx\ p_X(x)\ g(x), \text{ with } p_X(x) = \mathrm{Uniform}(a,b).$$

$$= \mathbb{E}[g(x)] \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i), \text{ with the } x_i \text{ drawn from } \mathrm{Uniform}(a,b).$$

Integral rewritten as an expectation value, and approximated as the sample mean.

More generally, $\int_a^b dx\ p_X(x)\ g(x) \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i)$, with x_i drawn from $p_X(x)$.

Quadrature as an expectation value (simple Monte Carlo)

Evaluate $\int_a^b dx\ g(x)$ using Monte Carlo sampling.

$$\int_a^b dx\ g(x) = (b-a)\int_a^b dx\ \frac{1}{b-a}\ g(x) = \int_a^b dx\ p_X(x)\ g(x), \text{ with } p_X(x) = \mathrm{Uniform}(a,b).$$

$$= \mathbb{E}[g(x)] \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i), \text{ with the } x_i \text{ drawn from } \mathrm{Uniform}(a,b).$$

Integral rewritten as an expectation value, and approximated as the sample mean.

More generally, $\int_a^b dx\ p_X(x)\ g(x) \approx \frac{1}{N}\sum_{i=1}^{N} g(x_i)$, with x_i drawn from $p_X(x)$.

Advantage: extremely easy to code/compute.

Disadvantage: Very slow convergence! Variance $\sim N^{-1}$, so error $\sim N^{-1/2}$.
Trapezoid and Simpson's Rule in d dimensions: error $\sim N^{-2/d}$ and $\sim N^{-4/d}$.
Monte Carlo methods become efficient in higher dimensions.

Example: Simple Monte Carlo vs. Rejection Sampling

Compute $\mathscr{I} = \int_{-1}^{1} dx\ e^{-x^2/2} = \sqrt{2\pi} \int_{-1}^{1} dx\ \varphi(x)$, where $\varphi(x)$ is the Standard Normal.

Exact value: $\mathscr{I} = \sqrt{2\pi}\ \Big(\Phi(1) - \Phi(-1)\Big)$; $\Phi(x) =$ CDF of the Standard Normal.

Example: Simple Monte Carlo vs. Rejection Sampling

Compute $\mathscr{I} = \int_{-1}^{1} dx\ e^{-x^2/2} = \sqrt{2\pi} \int_{-1}^{1} dx\ \varphi(x)$, where $\varphi(x)$ is the Standard Normal.

Exact value: $\mathscr{I} = \sqrt{2\pi}\ \Big(\Phi(1) - \Phi(-1)\Big)$; $\Phi(x) =$ CDF of the Standard Normal.

Method 1: choose $p_X(x) = \mathrm{Uniform}(-1, 1)$ and $g(x) = \varphi(x)$.

Example: Simple Monte Carlo vs. Rejection Sampling

Compute $\mathscr{I} = \int_{-1}^{1} dx\ e^{-x^2/2} = \sqrt{2\pi} \int_{-1}^{1} dx\ \varphi(x)$, where $\varphi(x)$ is the Standard Normal.

Exact value: $\mathscr{I} = \sqrt{2\pi}\ \left(\Phi(1) - \Phi(-1)\right)$; $\Phi(x) =$ CDF of the Standard Normal.

Method 1: choose $p_X(x) = \mathrm{Uniform}(-1, 1)$ and $g(x) = \varphi(x)$.

Method 2: choose $p_X(x) = \varphi(x)$ and $g(x) = \mathbb{I}_{x \in [-1,1]}(x)$.

Example: Simple Monte Carlo vs. Rejection Sampling

Compute $\mathscr{I} = \int_{-1}^{1} dx\ e^{-x^2/2} = \sqrt{2\pi} \int_{-1}^{1} dx\ \varphi(x)$, where $\varphi(x)$ is the Standard Normal.

Exact value: $\mathscr{I} = \sqrt{2\pi}\ \left(\Phi(1) - \Phi(-1)\right)$; $\Phi(x) =$ CDF of the Standard Normal.

Method 1: choose $p_X(x) = \mathrm{Uniform}(-1, 1)$ and $g(x) = \varphi(x)$.

Method 2: choose $p_X(x) = \varphi(x)$ and $g(x) = \mathbb{I}_{x\in[-1,1]}(x)$.

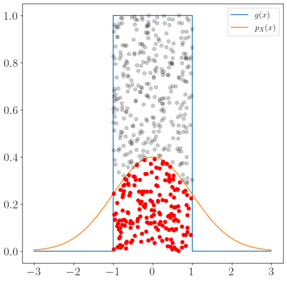
Method 3: rejection sampling.

Draw $X \sim \mathrm{Uniform}(-1, 1)$, $Y \sim \mathrm{Uniform}(0, 1)$.

$X \rightarrow$ range of x values for which $p_X(x)$ is desired.

$Y \rightarrow$ range of heights (values of $g(x)$).

Reject pairs with $y > p_X(x)$.



Example: Simple Monte Carlo vs. Rejection Sampling

Compute $\mathscr{I} = \int_{-1}^{1} dx\ e^{-x^2/2} = \sqrt{2\pi} \int_{-1}^{1} dx\ \varphi(x)$, where $\varphi(x)$ is the Standard Normal.

Exact value: $\mathscr{I} = \sqrt{2\pi}\ \big(\Phi(1) - \Phi(-1)\big)$; $\Phi(x) =$ CDF of the Standard Normal.

Method 1: choose $p_X(x) = \mathrm{Uniform}(-1, 1)$ and $g(x) = \varphi(x)$.

Method 2: choose $p_X(x) = \varphi(x)$ and $g(x) = \mathbb{I}_{x\in[-1,1]}(x)$.

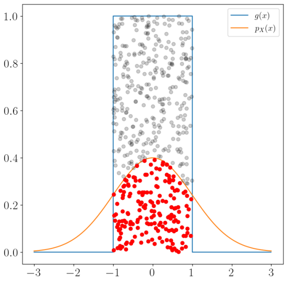
Method 3: rejection sampling.

Draw $X \sim \mathrm{Uniform}(-1, 1)$, $Y \sim \mathrm{Uniform}(0, 1)$.

$X \rightarrow$ range of x values for which $p_X(x)$ is desired.

$Y \rightarrow$ range of heights (values of $g(x)$).

Reject pairs with $y > p_X(x)$.



To the Jupyter notebook!

Problems with rejection sampling

Uses independent draws and is great for 1- or 2-dimensional problems.

For efficiency, need a good guess for the proposal distribution $g(x)$.

"Curse of dimensionality".

e.g., for N points along a dimension, the total # points required $\sim N^d$.

e.g., circle inscribed in a square, area ratio: $\frac{\pi}{4}$.

sphere inscribed in a cube, volume ratio: $\frac{\pi}{6}$.

d-dim. hypersphere/hypercube, hypervolume ratio: $\frac{\pi^{d/2}}{2^{d-1}\ d\ \Gamma(d/2)} \to 0$ as $d \to \infty$.

$P(\mathrm{rejection}) \uparrow$ as $d \uparrow$.

Need something better!

Importance sampling

Simple MC highly inefficient for precise estimates of rare events – give them larger weight so they are sampled more often.

Importance sampling

Simple MC highly inefficient for precise estimates of rare events – give them larger weight so they are sampled more often.

Sometimes easier to sample from a proposal distribution $q_X(x)$ instead of the target distribution $p_X(x)$:

$$\mathbb{E}[f(X)] = \int dx\ p_X(x)\ f(x) = \int dx q_X(x)\ \frac{p_X(x)}{q_X(x)}\ f(x) \equiv \int dx\ q_X(x)\ w(x) f(x),$$

where $w(x)$ is called the importance weight function.

Then, $\mathbb{E}_p[f(X)] = \mathbb{E}_q[w(X)f(X)] \approx \frac{1}{N}\sum_{i=1}^{N} w(X_i)f(X_i)$, where $X_i \sim q_X(x)$.

Compare to simple Monte Carlo method: $\mathbb{E}_p[f(X)] \approx \frac{1}{N}\sum_{i=1}^{N} f(X_i)$, where $X_i \sim p_X(x)$.

Importance sampling

Simple MC highly inefficient for precise estimates of rare events – give them larger weight so they are sampled more often.

Sometimes easier to sample from a proposal distribution $q_X(x)$ instead of the target distribution $p_X(x)$:

$$\mathbb{E}[f(X)] = \int dx\ p_X(x)\ f(x) = \int dx q_X(x)\ \frac{p_X(x)}{q_X(x)}\ f(x) \equiv \int dx\ q_X(x)\ w(x) f(x),$$

where $w(x)$ is called the importance weight function.

Then, $\mathbb{E}_p[f(X)] = \mathbb{E}_q[w(X)f(X)] \approx \frac{1}{N}\sum_{i=1}^{N} w(X_i)f(X_i)$, where $X_i \sim q_X(x)$.

Compare to simple Monte Carlo method: $\mathbb{E}_p[f(X)] \approx \frac{1}{N}\sum_{i=1}^{N} f(X_i)$, where $X_i \sim p_X(x)$.

Numerical considerations: for stability, w should be normalised, especially when one or both of $p_X(x)$ and $q_X(x)$ aren't.

The MC estimator for $\mathbb{E}_q[w(X)f(X)]$ is unbiased. For a smart choice of $q_X(x)$, it can also minimise variance.

In fact, one of the applications of importance sampling is to reduce the variance in MC estimates.

Importance sampling: example

See Jupyter notebook.

Some astronomical papers using importance sampling

Estimation of cosmological parameters:
Lewis & Bridle 2002, https://arxiv.org/abs/astro-ph/0205436
Trotta 2008, https://arxiv.org/abs/0803.4089

X-ray luminosity plane:
Gallo et al. (2018), http://adsabs.harvard.edu/abs/2018MNRAS.478L.132G

Extrasolar planet modelling:
Ford 2005, https://arxiv.org/abs/astro-ph/0512634.
Nelson et al. 2018, http://adsabs.harvard.edu/abs/2018arXiv180604683N.
Hsu et al. 2018, http://adsabs.harvard.edu/abs/2018AJ....155..205H.
Rajpaul et al. 2017, http://adsabs.harvard.edu/abs/2017MNRAS.471L.125R.