Advanced Simulation - Lecture 3

George Deligiannidis

January 27th, 2019

Transformation Method: pushforward

Let $\mathbb{Y}, \mathbb{X}$ be two topological spaces equipped with their Borel σ-algebras.

Suppose that $f : \mathbb{Y} \mapsto \mathbb{X}$ is Borel measurable;

Suppose that q is a Borel probability measure on $\mathbb{Y}$ and let $Y \sim q$.

Write π for the distribution of $X = f(Y)$, a Borel probability measure on $\mathbb{X}$.

Then π is the *push-forward of q under f*, written

$$\pi = f_* \mu.$$

It's defined as

$$\pi(B) = (f_*)\mu(B) = q\left(f^{-1}(B)\right), \quad \text{for all } B \in \mathcal{B}(\mathbb{X}).$$

In terms of expectations

$$\int h \circ \varphi \mathrm{d} q = \int h \mathrm{d} f_* \mu.$$

Transformation Method: change of variables formula

When $\mathrm{d}q(x) = q(x)\mathrm{d}x$, and φ is a bijection, then π also has a density given by the change of variables formula

$$\pi(x) = q \circ \varphi^{-1}(x) \left|\det(D\varphi^{-1})(x)\right| .$$

Transformation Method - Box-Muller Algorithm

Gaussian distribution. Let $U_1 \sim \mathcal{U}_{[0,1]}$ and $U_2 \sim \mathcal{U}_{[0,1]}$ be independent and set

$$R = \sqrt{-2\log\left(U_1\right)},\ \vartheta = 2\pi U_2.$$

Clearly R, ϑ independent and $R^2 \sim \mathrm{Exp}(1/2)$, $\vartheta \sim \mathcal{U}_{[0,2\pi]}$ with joint density

$$q(r^2, \vartheta) = \frac{1}{2\pi}\frac{1}{2}\exp(-r^2/2).$$

Set $X = R\cos(\vartheta)$, $Y = R\sin(\vartheta)$ a bijection.

Transformation Method - Box-Muller Algorithm

By standard facts:

$$\begin{aligned} f_{X,Y}(x,y) &= f_{R^2,\vartheta}(r^2(x,y),\theta(x,y))\Big| \det \frac{\partial(r^2,\vartheta)}{\partial(x,y)}\Big| \\ &= f_{R^2,\vartheta}(r^2(x,y),\theta(x,y))\Big| \det \frac{\partial(x,y)}{\partial(r^2,\vartheta)}\Big|^{-1} \\ &= \frac{1}{2}\frac{1}{2\pi}\exp\Big[-\frac{x^2+y^2}{2}\Big]2 = \frac{1}{2\pi}\exp\Big[-\frac{x^2+y^2}{2}\Big], \end{aligned}$$

since

$$\det \frac{\partial(x,y)}{\partial(r^2,\vartheta)}| = \begin{vmatrix} \frac{\cos(\vartheta)}{2r} & -r\sin\vartheta \\ \frac{\sin(\vartheta)}{2r} & r\cos\vartheta \end{vmatrix} = \frac{1}{2}.$$

thus (X,Y) are independent standard normal.

Transformation Method - Multivariate Normal

Let $Z = (Z_1, ..., Z_d) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$.
Let L be a real invertible $d \times d$ matrix satisfying $L\,L^T = \Sigma$, and $X = LZ + \mu$. Then $X \sim \mathcal{N}(\mu, \Sigma)$.

We have indeed $q(z) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2} z^T z\right)$ and

$$\pi(x) = q(z) \left|\det \partial z / \partial x\right|$$

where $\partial z/\partial x = L^{-1}$ and $\det\left(L^{-1}\right) = \det(\Sigma)^{-1/2}$. Additionally,

$$\begin{aligned} z^T z &= (x-\mu)^T \left(L^{-1}\right)^T L^{-1} (x-\mu) \\ &= (x-\mu)^T \Sigma^{-1} (x-\mu). \end{aligned}$$

In practice, use a Cholesky factorization $\Sigma = L\,L^T$ where L is a lower triangular matrix.

Sampling via Composition

Assume we have a joint pdf $\overline{\pi}$ with marginal π; i.e.

$$\pi(x) = \int \overline{\pi}_{X,Y}(x,y)\,dy$$

where $\overline{\pi}(x,y)$ can always be decomposed as

$$\overline{\pi}_{X,Y}(x,y) = \overline{\pi}_Y(y)\,\overline{\pi}_{X|Y}(x|\,y).$$

It might be easy to sample from $\overline{\pi}(x,y)$ whereas it is difficult/impossible to compute $\pi(x)$.

In this case, it is sufficient to sample

$$Y \sim \overline{\pi}_Y \text{ then } X|\,Y \sim \overline{\pi}_{X|Y}(\cdot|\,Y)$$

so $(X,Y) \sim \overline{\pi}_{X,Y}$ and hence $X \sim \pi$.

Latent variable models; HMMs;

Finite Mixture of Distributions

Assume one wants to sample from

$$\pi(x) = \sum_{i=1}^{p} \alpha_i.\pi_i(x)$$

where $\alpha_i > 0$, $\sum_{i=1}^{p} \alpha_i = 1$ and $\pi_i(x) \geq 0$, $\int \pi_i(x)\, dx = 1$.

We can introduce $Y \in \{1, ..., p\}$ and

$$\overline{\pi}_{X,Y}(x, y) = \alpha_y \times \pi_y(x).$$

To sample from $\pi(x)$, first sample Y from a discrete distribution such that $\mathbb{P}(Y = k) = \alpha_k$ then

$$X|(Y = y) \sim \pi_y.$$

Rejection Sampling

Basic idea: Sample from instrumental proposal $q \neq \pi$; correct through rejection step to obtain a sample from π.

Algorithm (Rejection Sampling). Given two densities π, q with $\pi(x) \leq M\, q(x)$ for all x, we can generate a sample from π by

1. Draw $X \sim q$, draw $U \sim \mathcal{U}_{[0,1]}$.
2. Accept $X = x$ as a sample from π if

$$U \leq \frac{\pi(x)}{M\, q(x)},$$

otherwise go to step 1.

Proposition

The distribution of the samples accepted by rejection sampling is π.

Rejection Sampling

Proof.

$$\mathbb{P}\left(X \in A \middle| X \text{ accepted}\right) = \frac{\mathbb{P}\left(X \in A, X \text{ accepted}\right)}{\mathbb{P}\left(X \text{ accepted}\right)}$$

where

$$\begin{aligned}
&\mathbb{P}\left(X \in A, X \text{ accepted}\right) \\
&= \int_{\mathbb{X}} \int_0^1 \mathbb{I}_A\left(x\right) \mathbb{I}\left(u \leq \frac{\pi\left(x\right)}{M\, q\left(x\right)}\right) q\left(x\right) du dx \\
&= \int_{\mathbb{X}} \mathbb{I}_A\left(x\right) \frac{\pi\left(x\right)}{M\, q\left(x\right)} q\left(x\right) dx \\
&= \int_{\mathbb{X}} \mathbb{I}_A\left(x\right) \frac{\pi\left(x\right)}{M} dx = \frac{\pi\left(A\right)}{M}.
\end{aligned}$$

□

Rejection Sampling

- Often we only know π and q up to some normalising constants; i.e.

$$\pi = \widetilde{\pi}/Z_\pi \quad \text{and} \quad q = \widetilde{q}/Z_q$$

where $\widetilde{\pi}, \widetilde{q}$ are known but Z_π, Z_q are unknown.
You still need to be able to sample from $q(\cdot)$.

- If you can upper bound:

$$\widetilde{\pi}(x)/\widetilde{q}(x) \leq \widetilde{M},$$

then using $\widetilde{\pi}$, $\widetilde{q}$ and $\widetilde{M}$ in the algorithm is correct.

- Indeed we have

$$\frac{\widetilde{\pi}(x)}{\widetilde{q}(x)} \leq \widetilde{M} \Leftrightarrow \frac{\pi(x)}{q(x)} \leq \widetilde{M}\frac{Z_q}{Z_\pi} = M.$$

Rejection Sampling

Let T denote the number of pairs (X, U) that have to be generated until X is accepted for the first time.

Lemma

T is geometrically distributed with parameter $1/M$ and in particular $\mathbb{E}(T) = M$.

In the unnormalised case, this yields

$$\mathbb{P}(X \text{ accepted}) = \frac{1}{M} = \frac{Z_\pi}{\widetilde{M} Z_q},$$

$$\mathbb{E}(T) = M = \frac{Z_q \widetilde{M}}{Z_\pi},$$

and it can be used to provide unbiased estimates of Z_π / Z_q and Z_q / Z_π.

Examples:Uniform from bounded subset of $\mathbb{R}^p$

- Let $B \subset \mathbb{R}^p$, a bounded subset of $\mathbb{R}^p$:

$$\pi(x) \propto \mathbb{I}_B(x).$$

  Let R be a rectangle containing $B \subset R$ and

$$q(x) \propto \mathbb{I}_R(x).$$

- Then we can use $\widetilde{M} = 1$ and

$$\widetilde{\pi}(x) / \left(\widetilde{M}' \widetilde{q}(x)\right) = \mathbb{I}_B(x).$$

- The probability of accepting a sample is then Z_π / Z_q.

Example: Normal density

- Let $\widetilde{\pi}(x) = \exp\left(-\frac{1}{2}x^2\right)$ and $\widetilde{q}(x) = 1/\left(1+x^2\right)$. We have

$$\frac{\widetilde{\pi}(x)}{\widetilde{q}(x)} = \left(1+x^2\right)\exp\left(-\frac{1}{2}x^2\right) \le 2/\sqrt{e} = \widetilde{M}$$

  which is attained at ± 1.
- Let $X \sim \widetilde{q}$. The acceptance probability is

$$\mathbb{P}\left(U \le \frac{\widetilde{\pi}(X)}{\widetilde{M}\widetilde{q}(X)}\right) = \frac{Z_\pi}{\widetilde{M}Z_q} = \frac{\sqrt{2\pi}}{\frac{2}{\sqrt{e}}\pi} = \sqrt{\frac{e}{2\pi}} \approx 0.66,$$

  and the mean number of trials to success is approximately $1/0.66 \approx 1.52$.

Examples: Genetic Linkage model

- We observe

$$(Y_1, Y_2, Y_3, Y_4) \sim \mathcal{M}\left(n; \frac{1}{2} + \frac{\theta}{4}, \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{\theta}{4}\right)$$

  where $\mathcal{M}$ is the multinomial distribution and $\theta \in (0,1)$.
- The likelihood of the observations is thus

$$\begin{aligned} & p(y_1, ..., y_4; \theta) \\ & = \frac{n!}{y_1! y_2! y_3! y_4!} \left(\frac{1}{2} + \frac{\theta}{4}\right)^{y_1} \left(\frac{1}{4}(1-\theta)\right)^{y_2+y_3} \left(\frac{\theta}{4}\right)^{y_4} \\ & \propto (2+\theta)^{y_1} (1-\theta)^{y_2+y_3} \theta^{y_4}. \end{aligned}$$

- Bayesian approach where we select $p(\theta) = \mathbb{I}_{[0,1]}(\theta)$ and are interested in

$$p(\theta | y_1, ..., y_4) \propto (2+\theta)^{y_1} (1-\theta)^{y_2+y_3} \theta^{y_4} \mathbb{I}_{[0,1]}(\theta).$$

Examples: Genetic linkage model

- Rejection sampling using the prior as proposal $q(\theta) = \widetilde{q}(\theta) = p(\theta)$ to sample from $p(\theta | y_1, ..., y_4)$.

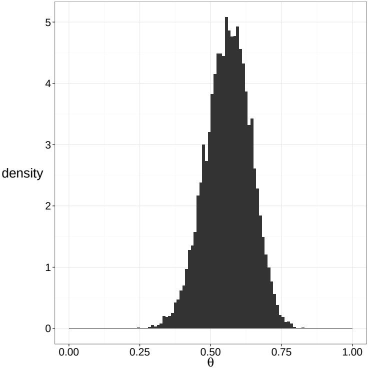
- To use accept-reject, we need to upper bound

$$\frac{\widetilde{\pi}(\theta)}{\widetilde{q}(\theta)} = \widetilde{\pi}(\theta) = (2+\theta)^{y_1} (1-\theta)^{y_2+y_3} \theta^{y_4}$$

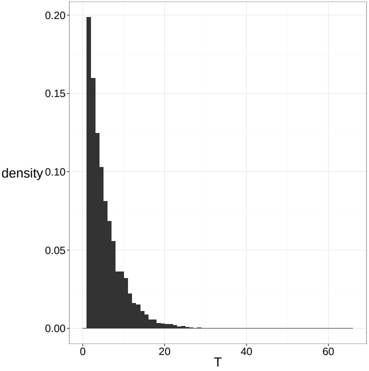
- Maximum of $\widetilde{\pi}$ can be found using standard optimization procedure to perform rejection sampling.

- For a realisation of (Y_1, Y_2, Y_3, Y_4) equal to $(69, 9, 11, 11)$ obtained with $n = 100$ and $\theta^{\star} = 0.6$, results shown in following figure.

Examples: Genetic linkage model



(a) Figure A



(b) Figure B

Figure: Histogram of 10,000 samples drawn from posterior obtained by rejection sampling (left); and histogram of waiting time distribution before acceptance (right).

Rejection Sampling Recap

Rejection sampling requires

- Samples from some distribution q;
- evaluation of $\pi(\cdot)$ point-wise, or unnormalized $\widetilde{\pi}$;
- an upper bound M on $\pi(x)/q(x)$, or $\widetilde{\pi}/q$ and so on.

Sometimes the upper bound is not feasible.

Importance Sampling

- We want to compute

$$I = \mathbb{E}_{\pi}(\varphi(X)) = \int_{\mathbb{X}} \varphi(x)\,\pi(x)\,dx.$$

- We do not know how to sample from the target π but have access to a proposal distribution of density q.

- We only require that

$$\pi(x) > 0 \Rightarrow q(x) > 0;$$

  i.e. the support of q includes the support of π.

- q is called the proposal, or importance distribution.

Importance Sampling

- We have the following identity

$$\boxed{I = \mathbb{E}_\pi(\varphi(X)) = \mathbb{E}_q(\varphi(X)w(X)),}$$

where $w : \mathbb{X} \to \mathbb{R}^+$ is the importance weight function

$$w(x) = \frac{\pi(x)}{q(x)}.$$

- Hence for $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} q$,

$$\widehat{I}_n^{\text{IS}} = \frac{1}{n} \sum_{i=1}^{n} \varphi(X_i) w(X_i).$$

Importance Sampling Properties

Proposition

(a) ***Unbiased:*** $\mathbb{E}_q[\widehat{I}_n^{IS}] = I$;

(b) ***Strongly consistent:*** *If* $\mathbb{E}_q(|\varphi(X)|\, w(X)) < \infty$ *then*

$$\lim_{n\to\infty} \widehat{I}_n^{IS} = I, \quad a.s.$$

(c) ***CLT:*** $\mathbb{V}_q(\widehat{I}_n^{IS}) = \sigma_{IS}^2/n$ *where*

$$\sigma_{IS}^2 := \mathbb{V}_q\left(\varphi(X) w(X)\right)$$

If $\sigma_{IS}^2 < \infty$ *then*

$$\lim_{n\to\infty} \sqrt{n}\left(\widehat{I}_n^{IS} - I\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma_{IS}^2\right).$$

Importance Sampling: Practical Advice

Consistency does not require $\sigma^2_{\mathsf{IS}} < \infty$ but highly recommended in practice (!).

Sufficient condition: If $\mathbb{E}_\pi\left(\varphi^2(X)\right) < \infty$ and $w(x) \leq M$ for all x for some $M < \infty$, then $\sigma^2_{\mathsf{IS}} < \infty$.

In practice ensure $w(x) \leq M$ although it is neither necessary nor sufficient, as seen in the following example.

Importance Sampling: Example

$$\pi\left(x\right)=\mathcal{N}\left(x;0,1\right),\ q\left(x\right)=\mathcal{N}\left(x;0,\sigma^{2}\right)$$

$$w(x)=\frac{\pi(x)}{q(x)}\propto\exp\left[-x^{2}\big(1-\frac{1}{\sigma^{2}}\big)\right].$$

For $\sigma^2 \geq 1$, $w\left(x\right) \leq M$ for all x,
and for $\sigma^2 < 1$, $w\left(x\right) \to \infty$ as $|x| \to \infty$.

For $\varphi\left(x\right) = x^2$, we have $\sigma^2_{\mathsf{IS}} < \infty$ for all $\sigma^2 > 1/2$.

For $\varphi\left(x\right) = \exp\left(\frac{\beta}{2}x^2\right)$, we have $I < \infty$ for $\beta < 1$
but $\sigma^2_{\mathsf{IS}} = \infty$ for $\beta > 1 - \frac{1}{2\sigma^2}$.

Optimal Importance Distribution I

Question

Is there a best proposal that minimizes the variance σ_{IS}^2?

Proposition

The optimal proposal minimising $\mathbb{V}_q\left(\widehat{I}_n^{IS}\right)$ is given by

$$q_{opt}(x) = \frac{|\varphi(x)|\,\pi(x)}{\int_{\mathbb{X}} |\varphi(x)|\,\pi(x)\,dx}.$$

Optimal Importance Distribution II

Proof.

We have indeed

$$\sigma_{\mathrm{IS}}^2 = \mathbb{V}_q\left(\varphi(X)w\left(X\right)\right) = \mathbb{E}_q\left(\varphi^2(X)w^2\left(X\right)\right) - I^2.$$

We also have by Jensen's inequality for any q

$$\mathbb{E}_q\left(\varphi^2(X)w^2\left(X\right)\right) \geq \left(\int_{\mathbb{X}} \left|\varphi(x)\right| \pi\left(x\right) dx\right)^2.$$

For $q = q_{\mathrm{opt}}$, we have

$$\begin{aligned}\mathbb{E}_{q_{\mathrm{opt}}}\left(\varphi^2(X)w^2\left(X\right)\right) &= \int_{\mathbb{X}} \frac{\varphi^2(x)\pi^2\left(x\right)}{\left|\varphi(x)\right| \pi\left(x\right)} dx \times \int_{\mathbb{X}} \left|\varphi(x)\right| \pi\left(x\right) dx \\ &= \left(\int_{\mathbb{X}} \left|\varphi(x)\right| \pi\left(x\right) dx\right)^2.\end{aligned}$$

□

Optimal Importance Distribution

$q_{\text{opt}}(x)$ can never be used in practice!

For $\varphi(x) > 0$ we have $q_{\text{opt}}(x) = \varphi(x)\pi(x)/I$ and $\mathbb{V}_{q_{\text{opt}}}\left(\widehat{I}_n^{\text{IS}}\right) = 0$ but this is because

$$\varphi(x)\, w(x) = \varphi(x)\, \frac{\pi(x)}{q_{\text{opt}}(x)} = I,$$

it requires knowing I!

This can be used as a guideline to select q; i.e. select $q(x)$ such that $q(x) \approx q_{\text{opt}}(x)$.

Particularly interesting in rare event simulation, not quite in statistics.

Normalised Importance Sampling

Standard IS has limited applications in statistics as it requires knowing $\pi(x)$ and $q(x)$ exactly.

Assume $\pi(x) = \widetilde{\pi}(x)/Z_\pi$ and $q(x) = \widetilde{q}(x)/Z_q$, $\pi(x) > 0 \Rightarrow q(x) > 0$ and and define

$$\widetilde{w}(x) = \frac{\widetilde{\pi}(x)}{\widetilde{q}(x)}.$$

An alternative identity is

$$\boxed{I = \mathbb{E}_\pi(\varphi(X)) = \frac{\int_{\mathbb{X}} \varphi(x)\, \widetilde{w}(x)\, q(x) dx}{\int_{\mathbb{X}} \widetilde{w}(x) q(x) dx}.}$$

SLLN for NIS

Proposition (SLLN for NIS)

Let $X_1, ..., X_n \overset{i.i.d.}{\sim} q$ and assume that $\mathbb{E}_q(|\varphi(X)|\, w(X)) < \infty$. Then

$$\widehat{I}_n^{NIS} = \frac{\sum_{i=1}^n \varphi(X_i)\widetilde{w}(X_i)}{\sum_{i=1}^n \widetilde{w}(X_i)}$$

is strongly consistent.

Proof.

Divide numerator and denominator by n. Both converge almost surely by the strong law of large numbers. □

BUT, for finite n $\widehat{I}_n^{\text{NIS}}$ is **biased**, see notes Chapter 3.

CLT for NIS

Proposition

If $\mathbb{V}_q(\varphi(X)w(X)) < \infty$ *and* $\mathbb{V}_q(w(X)) < \infty$ *then*

$$\sqrt{n}(\widehat{I}_n^{NIS} - I) \Rightarrow \mathcal{N}(0, \sigma^2_{NIS}),$$

where

$$\begin{aligned}\sigma^2_{NIS} &:= \mathbb{V}_q\Big(\big[\varphi(X)w(X)) - Iw(X)\big]\Big)\\ &= \int \frac{\pi(x)^2\,(\varphi(x) - I)^2}{q(x)}\mathrm{d}x.\end{aligned}$$

Proof

Proof.

First notice that with $X_1, \ldots, X_n$ i.i.d. $\sim q$

$$\sqrt{n}(\widehat{I}_n^{\text{NIS}} - I) = \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n \widetilde{w}(X_i)\big[\varphi(X_i) - I\big]}{\frac{1}{n} \sum_{i=1}^n \widetilde{w}(X_i)}$$

where since $\widetilde{w}(x) = \widetilde{\pi}/\widetilde{q}$

$$\mathbb{E}_q\Big[\widetilde{w}(X_n)(\varphi(X_i) - I)\Big] = 0.$$

Since $\mathbb{V}_q(\varphi(X)w(X)) < \infty$ by standard CLT

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \widetilde{w}(X_i)\big[\varphi(X_i) - I\big] \Rightarrow \mathcal{N}\Big(0, \mathbb{V}_q\Big(\widetilde{w}(X_1)[\varphi(X_1) - I]\Big)\Big).$$

Proof ctd...

Proof.

The strong law of large numbers applied to the denominator

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{w}(X_i) \to \mathbb{E}_q[\widetilde{w}(X_1)] = Z_\pi/Z_q, \quad \text{a.s.}$$

By Slutsky's theorem, combining the two

$$\sqrt{n}(\widehat{I}_n^{\text{NIS}} - I) \Rightarrow \mathcal{N}\Big(0, \mathbb{V}_q\big(\widetilde{w}(X_1)[\varphi(X_1) - I]\big)\frac{Z_q^2}{Z_\pi^2}\Big)$$
$$\sim \mathcal{N}\Big(0, \sigma^2_{\text{NIS}}\Big).$$

$\square$

Alternatively, use Delta method.

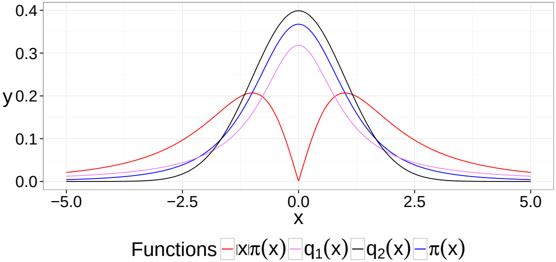
Toy Example: t-distribution

- We want to compute $I = \mathbb{E}_\pi(|X|)$ where $\pi(x) \propto \left(1 + x^2/3\right)^{-2}$ (t_3-distribution).

(a) Directly sample from π.

(b) Use $q_1(x) = g_{t_1}(x) \propto \left(1 + x^2\right)^{-1}$ (t_1-distribution).

(c) Use $q_2(x) \propto \exp\left(-x^2/2\right)$ (normal).



Toy Example: t-distribution

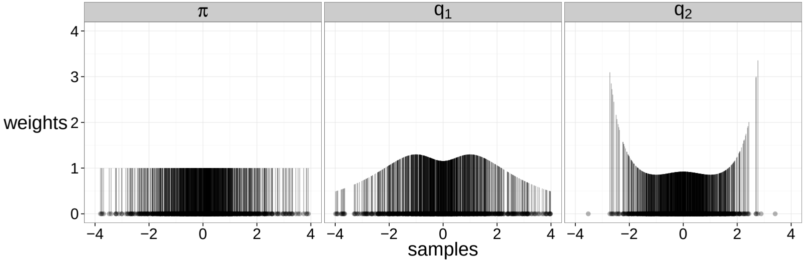


Figure: Sample weights obtained for 1000 realisations of X_i, from the different proposal distributions.

Toy Example: t-distribution

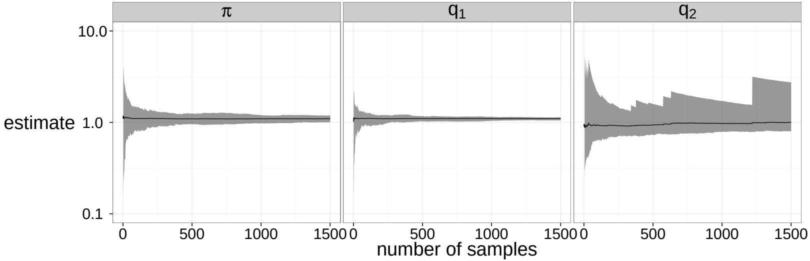


Figure: Estimates $\widehat{I_n}$ of I obtained after 1 to 1500 samples. The grey shaded areas correpond to the range of 100 independent replications.