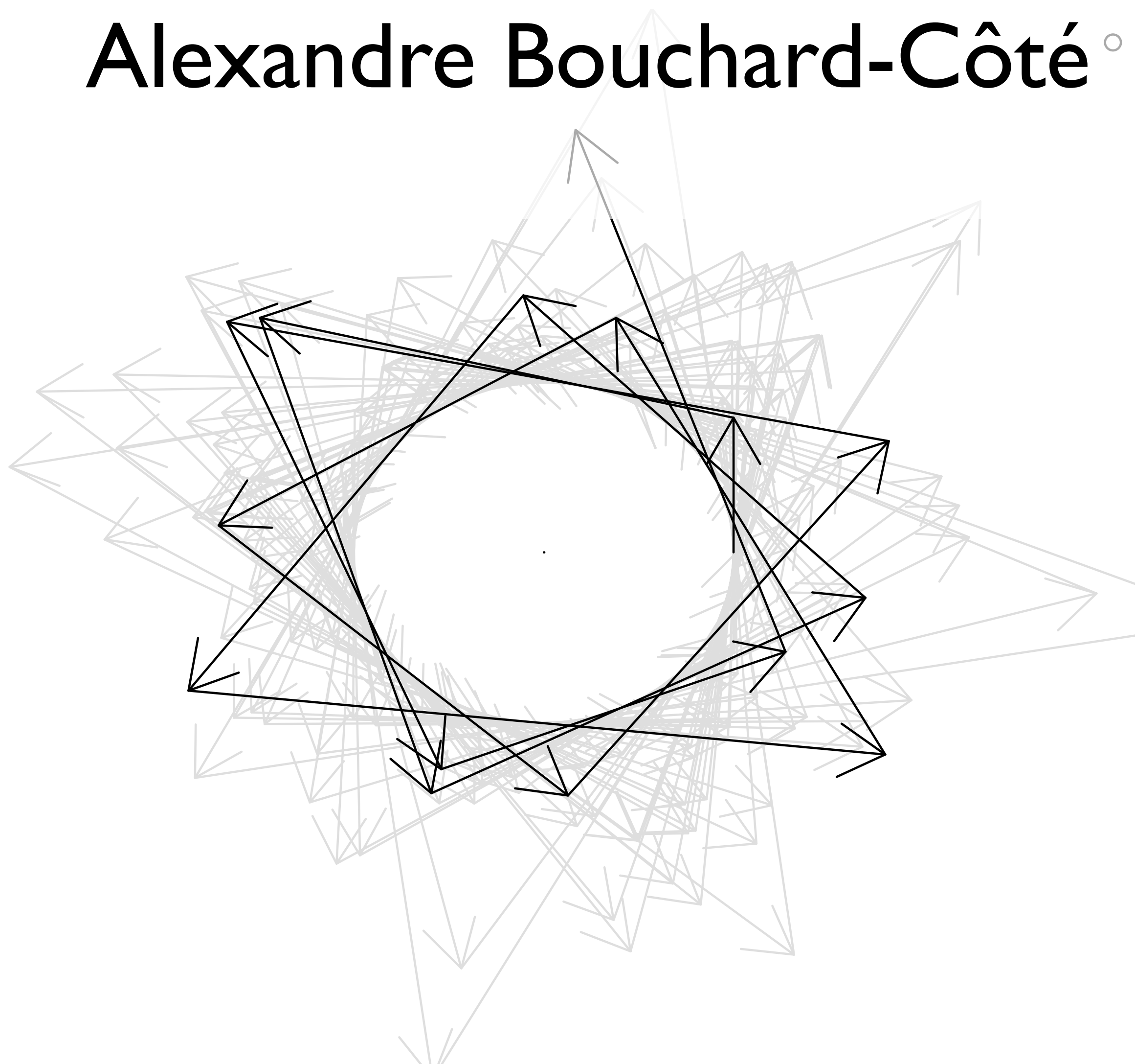
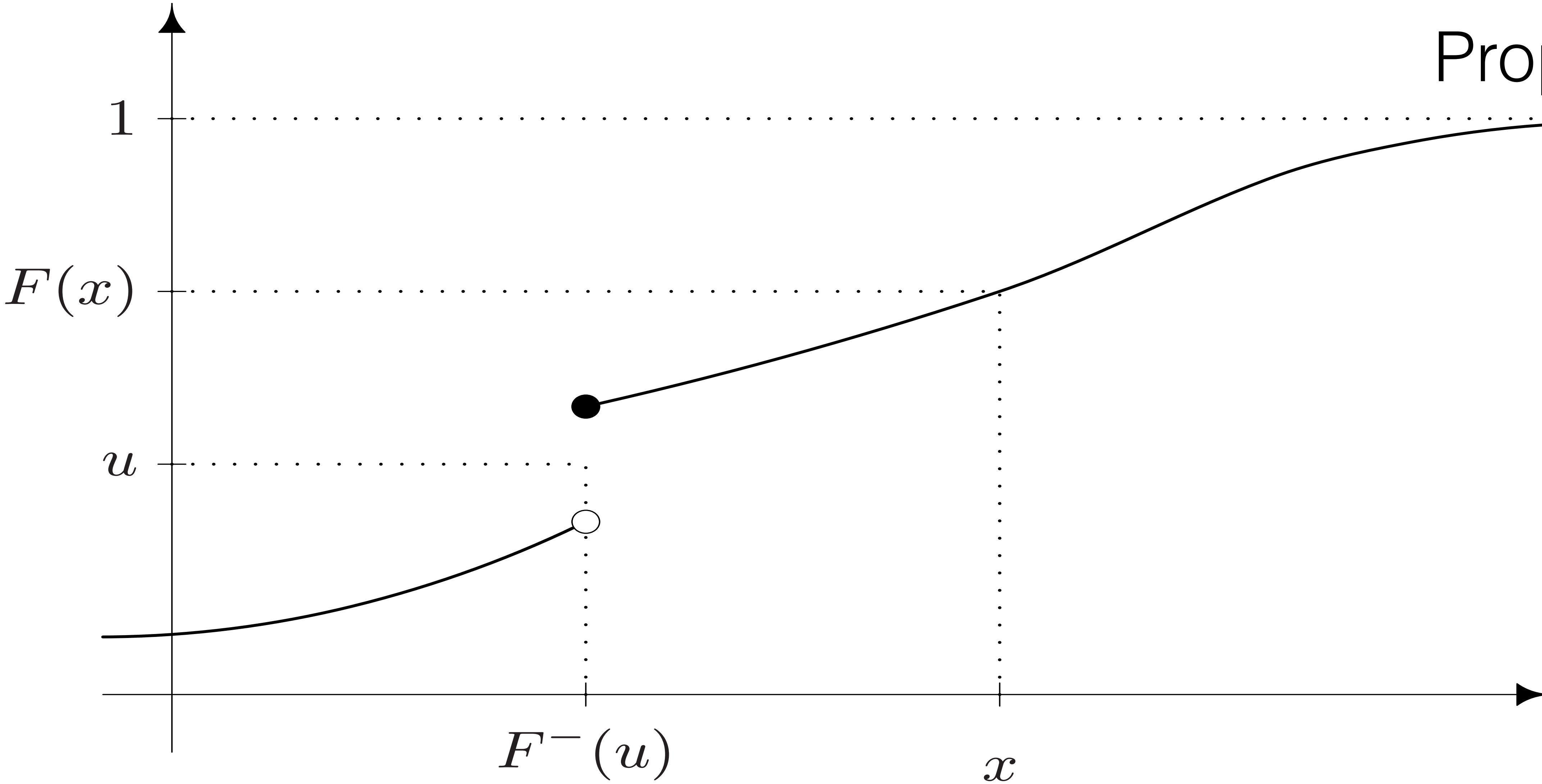
Monte Carlo methods

Alexandre Bouchard-Côté



Review / Q&A / Exercise solutions



- **Proposition**. Let F be a cdf and $U \sim \mathcal{U}_{[0,1]}$. Then $X = F^{-}(U)$ has cdf F.

Proof. $F^{-}(u) \leq x \Leftrightarrow u \leq F(x)$ so $U \sim \mathcal{U}_{[0,1]}$, we have

$$\mathbb{P}\left(F^{-}(U) \leq x\right) = \mathbb{P}\left(U \leq F(x)\right) = F(x).$$

Exercise: construct a RNG for exponential random variables of a given rate

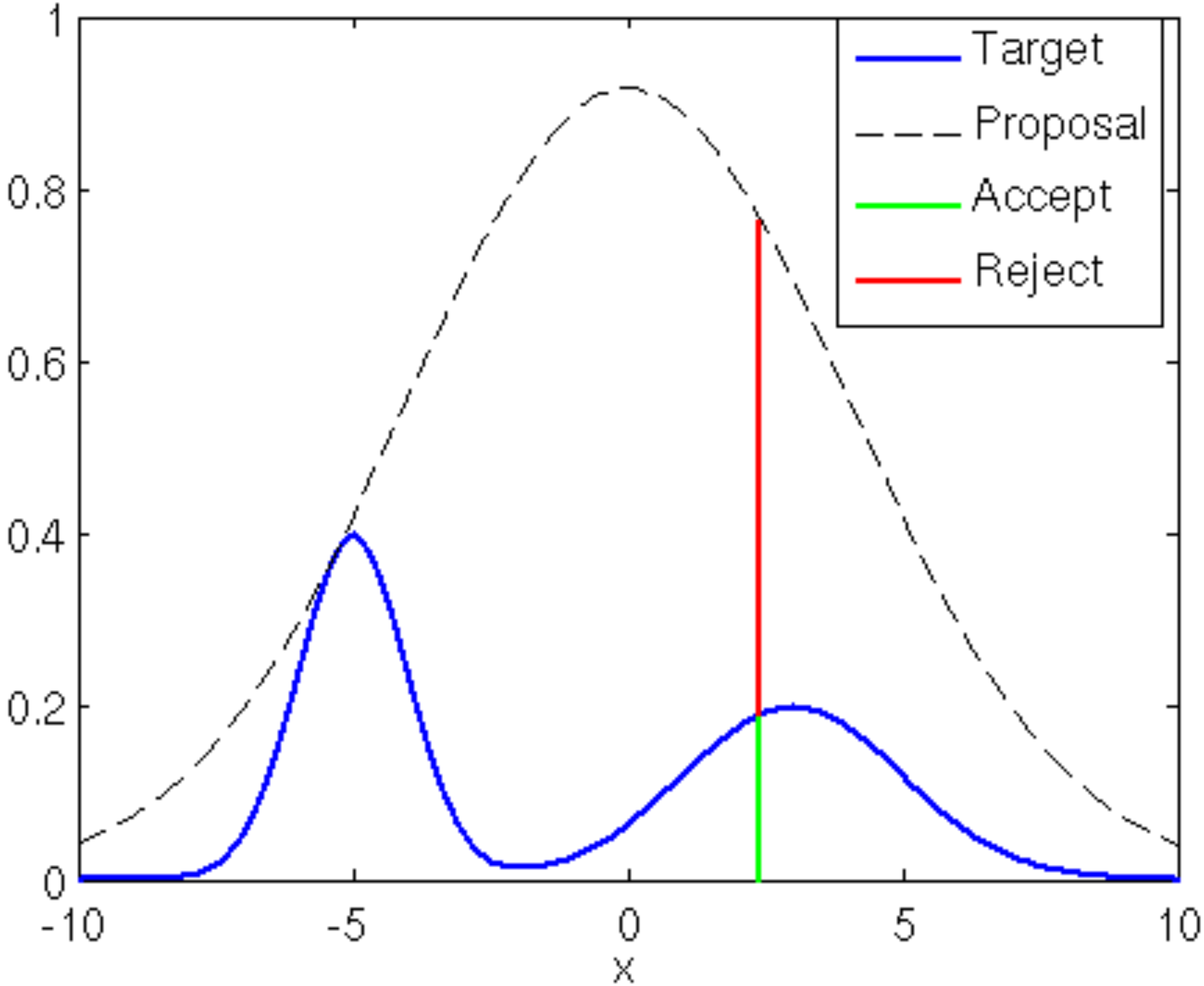
Transformation method

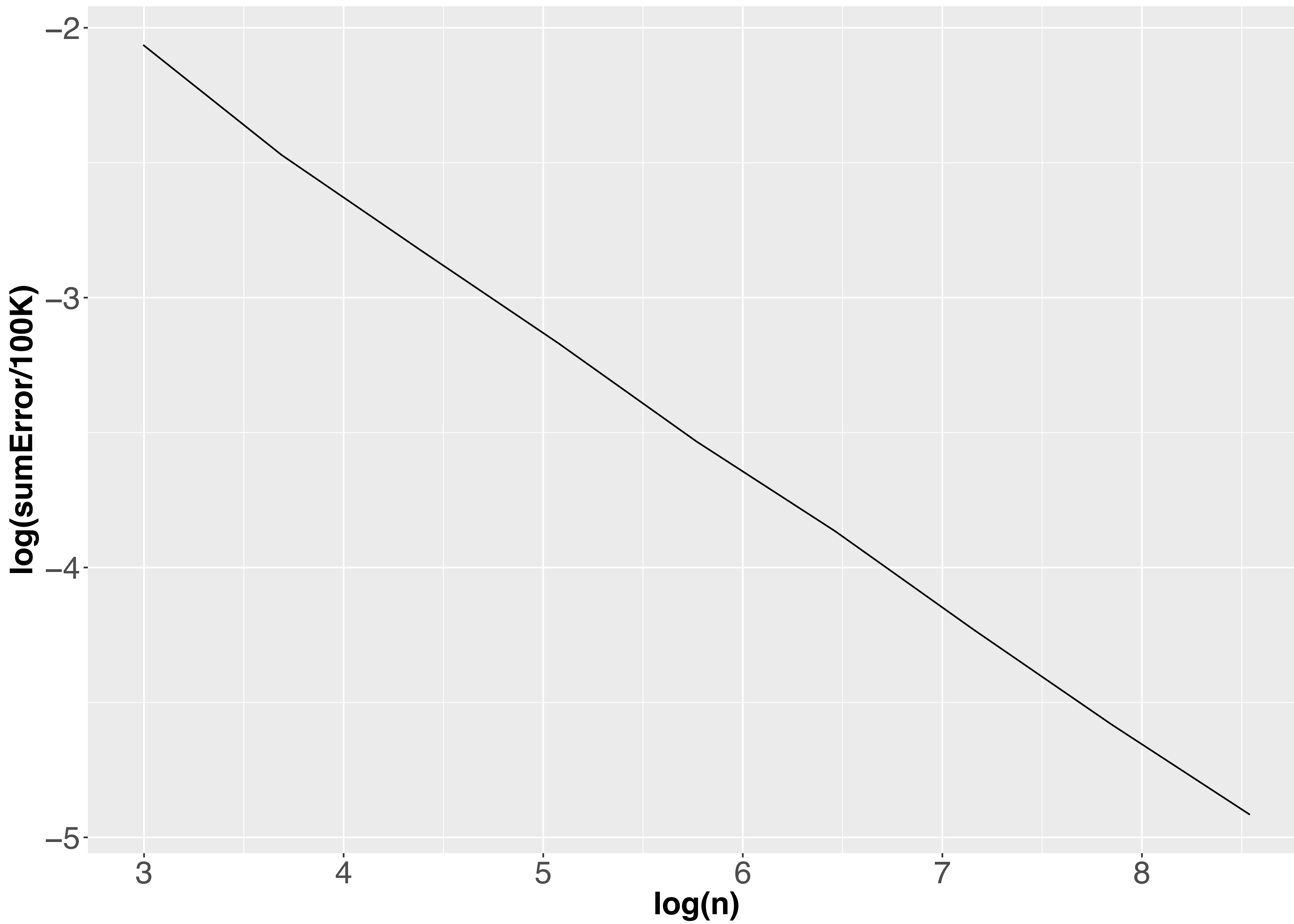
- Let $Y \sim q$ be a $\mathbb{Y}$-valued random variable (rv) of, which we can simulate (eg, by inversion)
- Let $X \sim \pi$ be a $\mathbb{X}$-valued rv, which we wish to simulate.
- It may be that we can find a function $\varphi : \mathbb{Y} \to \mathbb{X}$ with the property that if we simulate $Y \sim q$ and then set $X = \varphi(Y)$ then we get $X \sim \pi$.
- Inversion is a special case of this idea.

Augmentation / auxiliary variables

Def. 16

- Example: how to simulate from a mixture model?
- **Key idea**: marginalization is easy with Monte Carlo methods
  - Contrast with analytical marginalization
- Build (X,Y) such that distribution of interest is a marginal (or a conditional on a discrete variable)
- **Exercises**:
  - write pseudo-code to simulate from a mixture distribution with 2 normal components,
  - show that rejection sampling is an augmentation sampling scheme


Target
Proposal
Accept
Reject
1
0.8
0.6
0.4
0.2
0
-10
-5
0
5
10
X



1. Use this plot to empirically derive the running time of Simple Monte Carlo for a given tolerance *tol*. Create the plot for Example 7 [c.f. lecture slides].

4. Compute the running time in *tol* and *d* for Example 7 [c.f. lecture slides] but with non-diagonal covariance normal vectors.

Fundamental equation to analyze Monte Carlo methods

running time = [number of samples needed to get a tolerance (with probability 95%)] x [compute cost per sample]

- Exercise:
  - Compute the running time in *tol* and *d* for Example 7 but with non-diagonal covariance normal vectors

Using CLT in practice

- Let X be a random tree
- We are looking at a clade indicator $f(X)$ as in Example 2.
- After 500 iid trees your MC estimate for the clade support is roughly 10%
- Should you extract more samples?
  - Say we want a scheme with relative error of less than 10% for approximately 95% of the random seeds

Importance sampling (IS)

IS: when proposal and target have known normalizations

If: $\pi(x) > 0 \Rightarrow q(x) > 0$

Then:

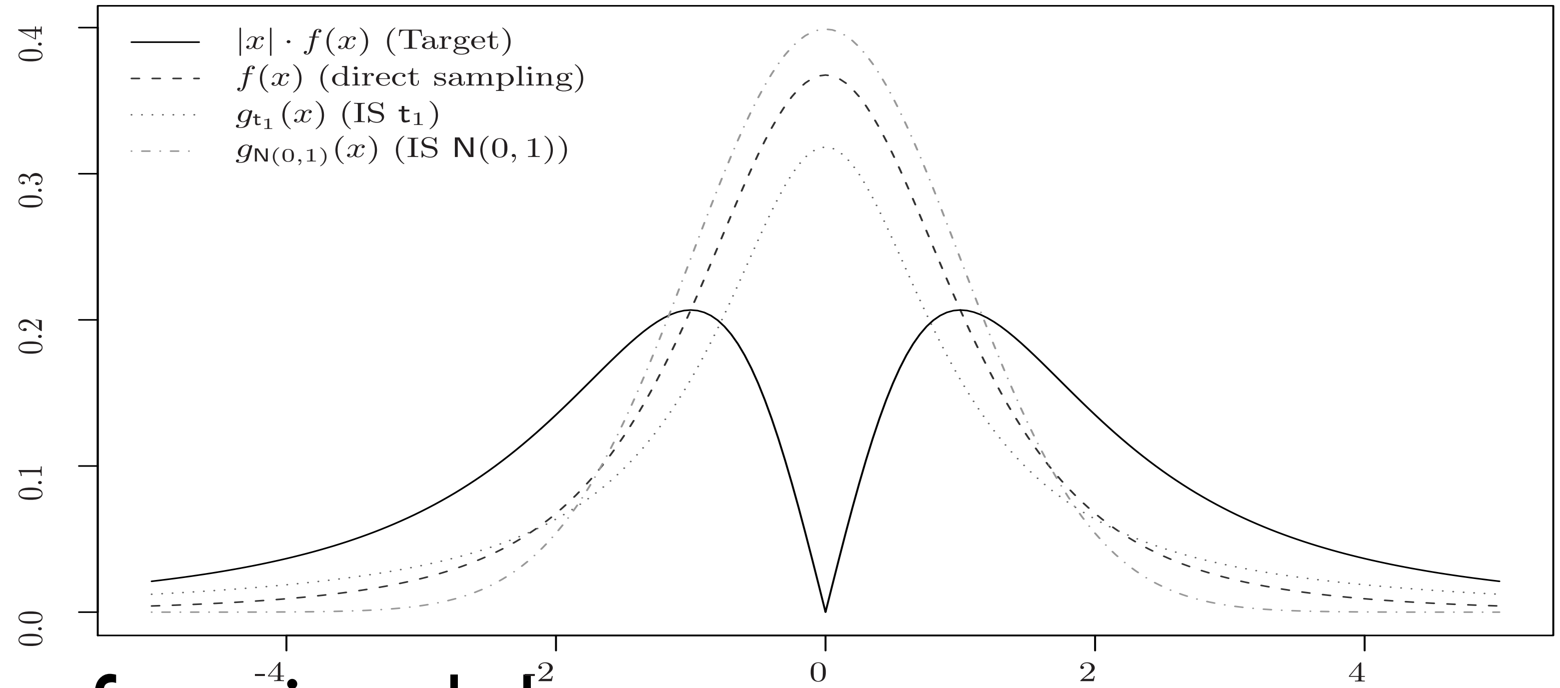
$$\begin{aligned}\pi(A) &:= \int_A \pi(x)\,dx \\ &= \int_A \underbrace{\frac{\pi(x)}{q(x)}}_{:=w(x)} q(x)\,dx \\ &= \int_A w(x)\,q(x)\,dx\end{aligned}$$

$X_1, ..., X_n$ be a sample of independent random variables distributed according to q

$$\widehat{I}_n^{\mathrm{IS}} = \frac{1}{n}\sum_{i=1}^{n} \phi(X_i) w(X_i)$$

Example/exercise

Ex 19



- Test function: |x|
- Target density: t-distribution, 3 degrees of freedom
- Compare (x-axis, 1-1500, y axis, partial sum, range of 100 replicates)
  - Simple MC
  - IS with t proposal, 1 degree of freedom
  - IS with normal proposal

Normalization: important issue in large scale Bayesian or random effects models

- We are given a density known up to a normalization constant

$$\pi(x) = \frac{\gamma(x)}{Z}$$

- Example:

$$\pi(x) = \frac{\text{joint}(x, y)}{\text{evidence}(y)}$$

x : unknown
y : data

- We want a law of large number

$$\frac{1}{N}\sum_{i=1}^{N}\varphi(X^{(i)}) \to \int \varphi(x)\pi(\mathrm{d}x) \text{ a.s.}$$

'test function'

Note: *We almost never care about the samples themselves!*

IS: when proposal and target *do not* have known normalizations

If: $\pi(x) > 0 \Rightarrow q(x) > 0$

Then:
$$I = \mathbb{E}_\pi(\phi(X)) = \int_{\mathbb{X}} \phi(x)\,\pi(x)\,dx$$
$$= \frac{\int_{\mathbb{X}} \phi(x)\,w(x)\,q(x)\,dx}{\int_{\mathbb{X}} w(x)\,q(x)\,dx}$$
$$= \frac{\mathbb{E}_q(\phi(X)w(X))}{\mathbb{E}_q(w(X))}.$$

$X_1, ..., X_n$ be a sample of independent random variables distributed according to q

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

Analysis of the asymptotic variance

Assume that $\mathbb{V}_q\left(\phi(X)w\left(X\right)\right) < \infty$ and $\mathbb{V}_q\left(w\left(X\right)\right) < \infty$ then

$$\sqrt{n}\left(\widehat{I}_n^{\mathrm{NIS}} - I\right) \xrightarrow{\mathrm{D}} \mathcal{N}\left(0, \sigma^2_{\mathrm{NIS}}\right)$$

Exercise: compute asymptotic variance

Tool: delta method

If: $$\sqrt{n}\left(Z_n - \mu\right) \xrightarrow{D} \mathcal{N}\left(0, \Sigma\right).$$

Then: $$\sqrt{n}\left(g\left(Z_n\right) - g\left(\mu\right)\right) \rightarrow \mathcal{N}\left(0, \nabla^T g\left(\mu\right)\ \Sigma\ \nabla g\left(\mu\right)\right).$$