Bayesian Learning

Video Lecture 0 - Monte Carlo methods

Mattias Villani

Department of Statistics
Stockholm University

mattiasvillani.com @matvil @matvil mattiasvillani

Monte Carlo simulation

- Interest: **expectations of functions** $f(X)$

$$\mathbb{E}_p(f(X)) = \begin{cases} \sum_{\text{all } x} f(x) \cdot p(x) & \text{for discrete } X \\ \int f(x) \cdot p(x)\, \mathrm{d}x & \text{for continuous } X \end{cases}$$

- Examples:
  - **Mean** $\mathbb{E}(X)$

$$f(x) = x$$

  - **Variance** $\mathbb{V}(X)$

$$f(x) = (x - \mu)^2$$

  - **Probabilities** $\Pr(a \leq X \leq b)$

$$f(x) = \begin{cases} 1 & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Properties of sample averages

- **Law of Large Numbers (LLN)**: sample means $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ converges to population means $\mu = \mathbb{E}_p(X)$

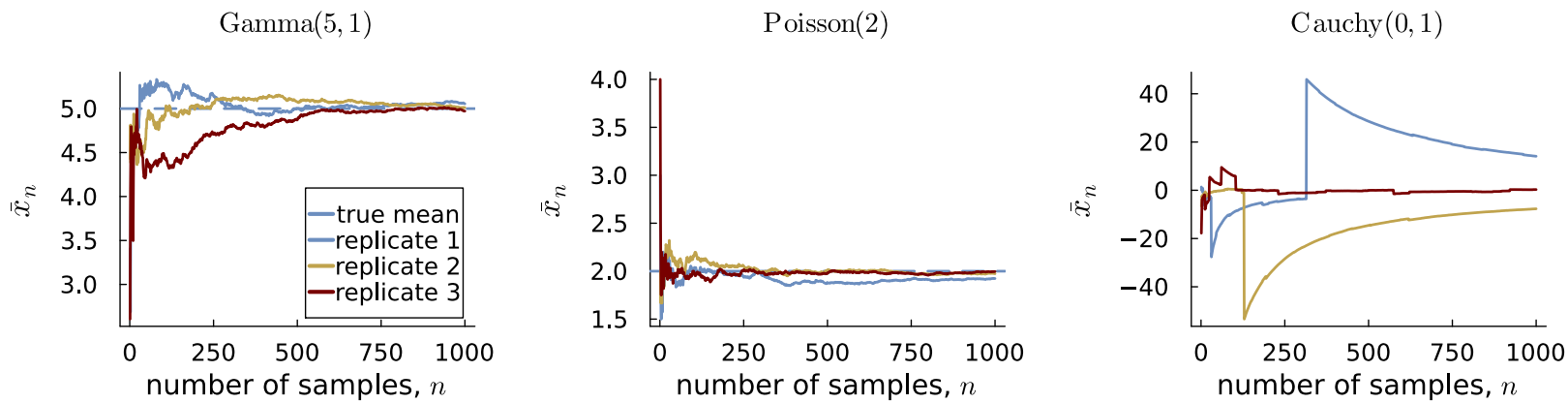
$$\bar{X}_n \xrightarrow{p} \mathbb{E}_p(X)$$

where $X_1, \ldots, X_n$ are iid from $p(x)$.

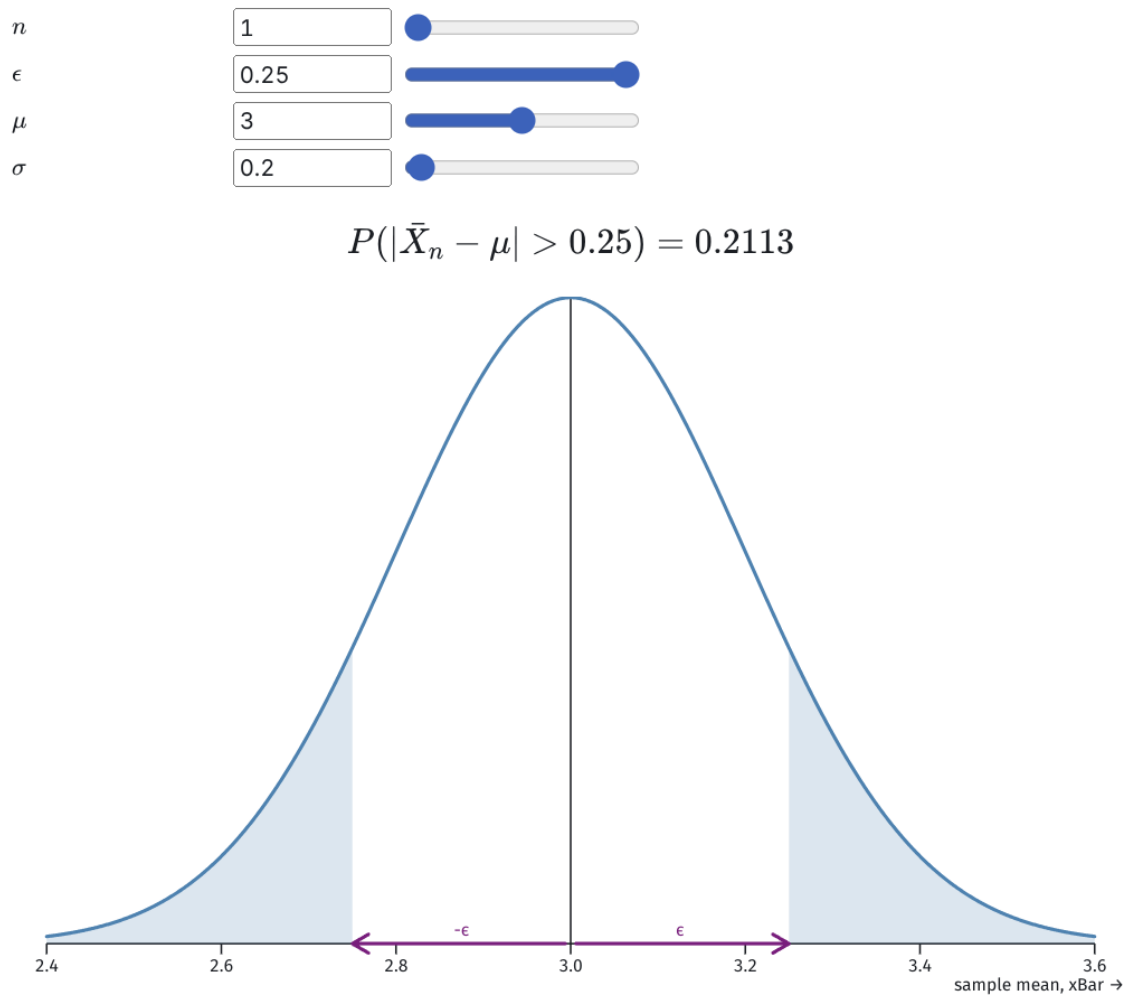
- **Central limit theorem (CLT)**: informally, for large n

$$\bar{X}_n \overset{\text{approx}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right),$$

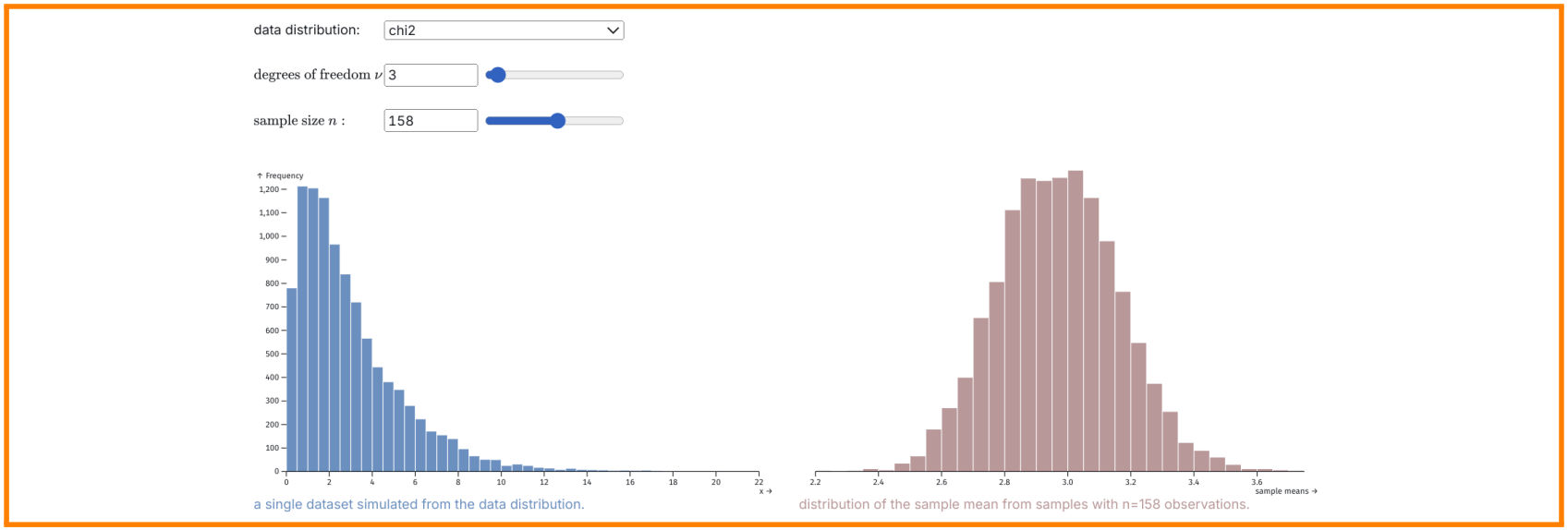
where $\sigma^2 = \mathbb{V}(X)$.



Law of large numbers - widget



Central limit theorem - widget



Sample averages of functions

- **LLN for a function $f(x)$**

$$\bar{f}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow{p} \mathbb{E}_p\left(f(X)\right)$$

where $X_1, \ldots, X_n$ are iid from $p(x)$.

- **CLT for a function $f(x)$**

$$\bar{f}_n \overset{\text{approx}}{\sim} N\left(\mu_f, \frac{\sigma_f^2}{n}\right),$$

where $\mu_f = \mathbb{E}_p\left(f(X)\right)$ and $\sigma_f^2 = \mathbb{V}_p\left(f(X)\right)$.

- We can estimate σ_f^2 by sample variance of function evaluations

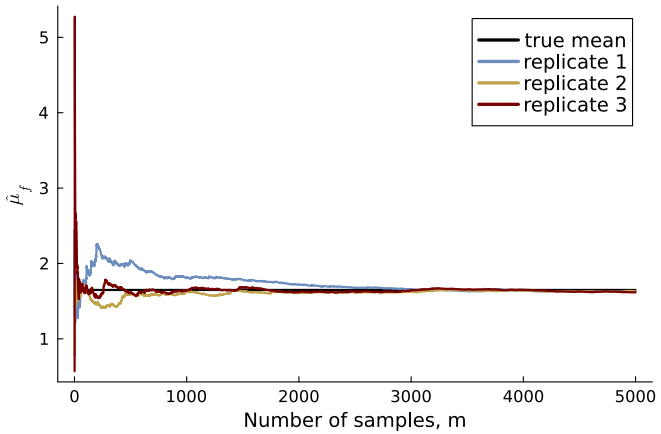
$$\hat{\sigma}_f^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(f(x_i) - \bar{f}_n\right)^2$$

Monte Carlo simulation

Monte Carlo integration

$$\hat{\mu}_f = \frac{1}{m}\sum_{i=1}^{m} f(x_i)\,, \qquad x_i \overset{\text{iid}}{\sim} p(x)$$

- **Unbiased** $\mathbb{E}_p(\hat{\mu}_f) = \mu_f$ for all m.
- **LLN**: **consistent** $\hat{\mu}_f \xrightarrow{p} \mu_f$ as $m \to \infty$.
- **CLT**: quantifies **estimation error** $\hat{\mu}_f - \mu_f$ for finite m.
- Example: $f(x) = \exp(x)$ and $X \sim N(0,1)$.
  True mean: $\mathbb{E}_p(\exp(X)) = \exp(1/2) \approx 1.648$



Monte Carlo simulation - multivariate

- Two extensions:
  - Multivariate input: $\boldsymbol{x}$
  - Multivariate output: $\boldsymbol{f}(\boldsymbol{x})$
- Example: $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$.
- **Monte Carlo integration**

$$\hat{\boldsymbol{\mu}}_f = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{f}(\boldsymbol{x}_i), \qquad \boldsymbol{x}_i \overset{\text{iid}}{\sim} p(\boldsymbol{x})$$

- Still unbiased and consistent.
- Multivariate **CLT**:

$$\hat{\boldsymbol{\mu}}_f \overset{\text{approx}}{\sim} N\left(\boldsymbol{\mu}_f, \frac{1}{m} \mathbb{V}_p\left(\boldsymbol{f}(X)\right)\right),$$

  where $\mathbb{V}_p\left(\boldsymbol{f}(X)\right)$ is a $d \times d$ covariance matrix.
- Same $\frac{1}{m}$ rate in multivariate case. Works in high dimensions!

Importance sampling

- Aim:

$$\mathbb{E}_p\left(f(X)\right) = \int f(x)\, p\left(x\right) \mathrm{d}x$$

- **Importance sampling** (**IS**) samples from a **proposal distribution** $q\left(x\right)$ instead of target distribution $p\left(x\right)$.
- Based on the trivial identity

$$\int f(x)\, p\left(x\right) \mathrm{d}x = \int f(x)\, \frac{p\left(x\right)}{q\left(x\right)} q\left(x\right) \mathrm{d}x = \int f(x)\, w\left(x\right)\, q(x) \mathrm{d}x$$

- **Weight function** (importance function)

$$w\left(x\right) = \frac{p\left(x\right)}{q\left(x\right)}$$

- **Importance sampling estimator**

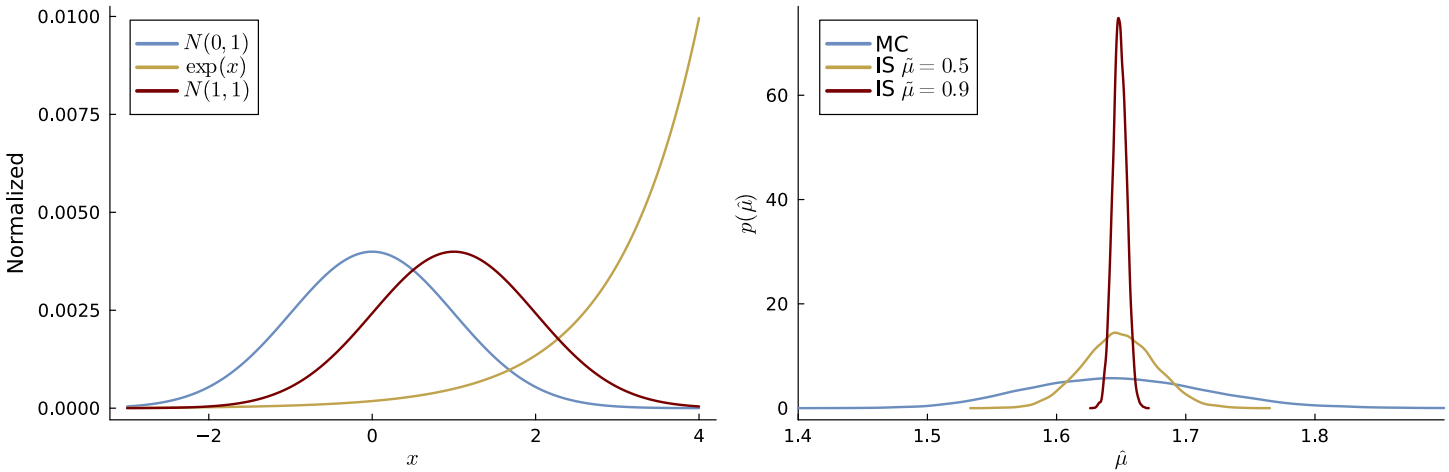
$$\hat{\mu}_f^{\mathrm{IS}} = \frac{1}{m} \sum_{i=1}^{m} f(x_i)\, w(x_i), \qquad x_i \overset{\mathrm{iid}}{\sim} q(x)$$

Importance sampling

- **Importance sampling estimator**

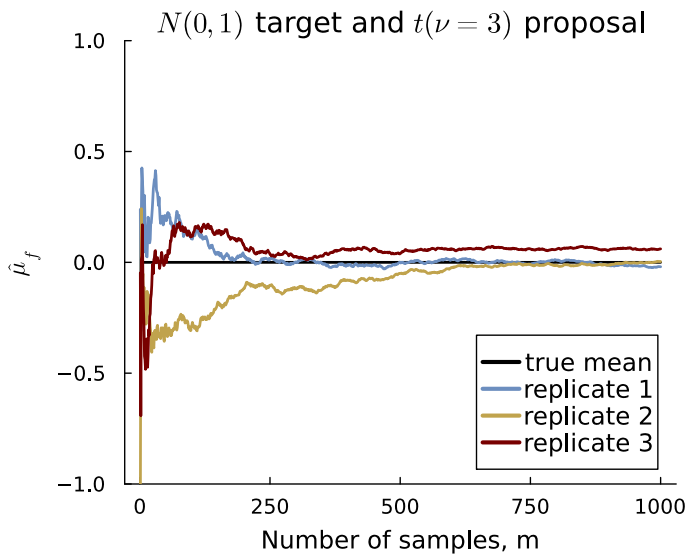
$$\hat{\mu}_f^{\mathrm{IS}} = \frac{1}{m}\sum_{i=1}^{m} f(x_i)\, w(x_i), \qquad x_i \overset{\mathrm{iid}}{\sim} q(x)$$

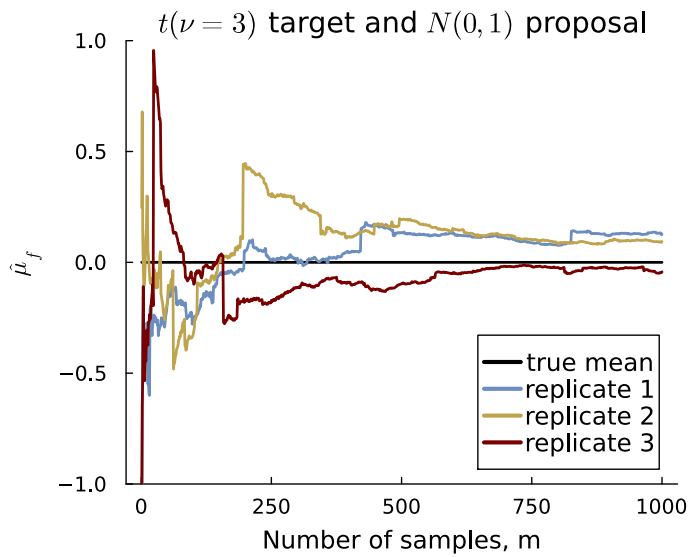
- **Unbiased** $\mathbb{E}_q(\hat{\mu}_f^{\mathrm{IS}}) = \mu_f$ and **consistent** $\hat{\mu}_f^{\mathrm{IS}} \xrightarrow{p} \mu_f$.
- The proposal distribution determines the efficiency (variance).
- Proposal $q(x)$ proportional to integrand $f(x)\,p(x)$ is ideal.
- Example: $f(x) = \exp(x)$ and $X \sim N(0,1)$. Proposal: $N(\tilde{\mu}, 1)$.



The tails of the proposal are important

- **Proposal** $q(x)$ should have **heavier tails** than target $p(x)$.
- Example 1: (OK!):
  - Target: $N(0, 1)$
  - Proposal: Student-t with three df.
- Example 2 (not OK!):
  - Target: Student-t with three df.
  - Proposal: $N(0, 1)$





Rejection sampling

- Rejection sampling samples from the target distribution $p(x)$.
- Draws from proposal $q(x) \neq p(x)$ and then accepts/rejects draws.
- Need to find a majorization constant

$$p(x) \leq M \cdot q(x) \qquad \text{for all } x$$

Rejection sampling:

1. Draw a candidate $x^\star$ from the proposal distribution $q(x)$.
2. Draw $u \sim \text{Uniform}(0, 1)$.
3. Accept the candidate $x^\star$ if

$$u < \frac{p(x^\star)}{M \cdot q(x^\star)}$$

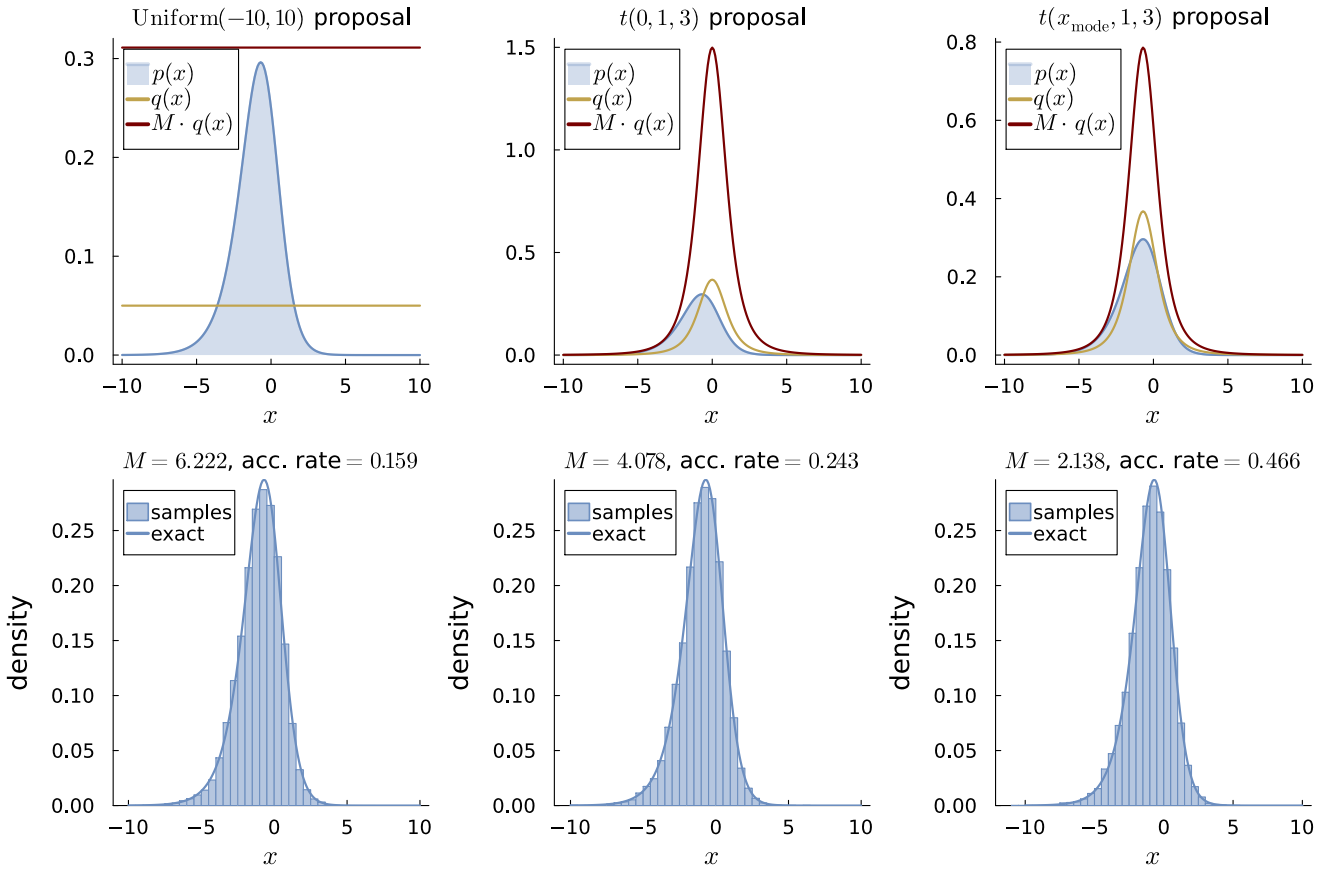
Otherwise, reject the candidate and return to step 1.

Rejection sampling - example

- Simulation from Logistic-Beta density

$$p(x) = \frac{\sigma(x)^\alpha \sigma(-x)^\beta}{B(\alpha, \beta)}$$

where $\sigma(x) = 1/(1 + e^{-x})$ is the logistic function.



Unnormalized target density

- Both Importance sampling and Rejection sampling works with **unnormalized target density**

$$p(x) = \frac{\tilde{p}(x)}{\int \tilde{p}(x)dx}$$

- The unnormalized density $\tilde{p}(x)$ is known ...
- ... but **normalizing constant** $\int \tilde{p}(x)dx$ cannot be computed.
- Common in **Bayesian learning** where the target is the **posterior distribution**

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)d\theta}$$

- (Much) more later!