Lecture 6: Monte Carlo Methods

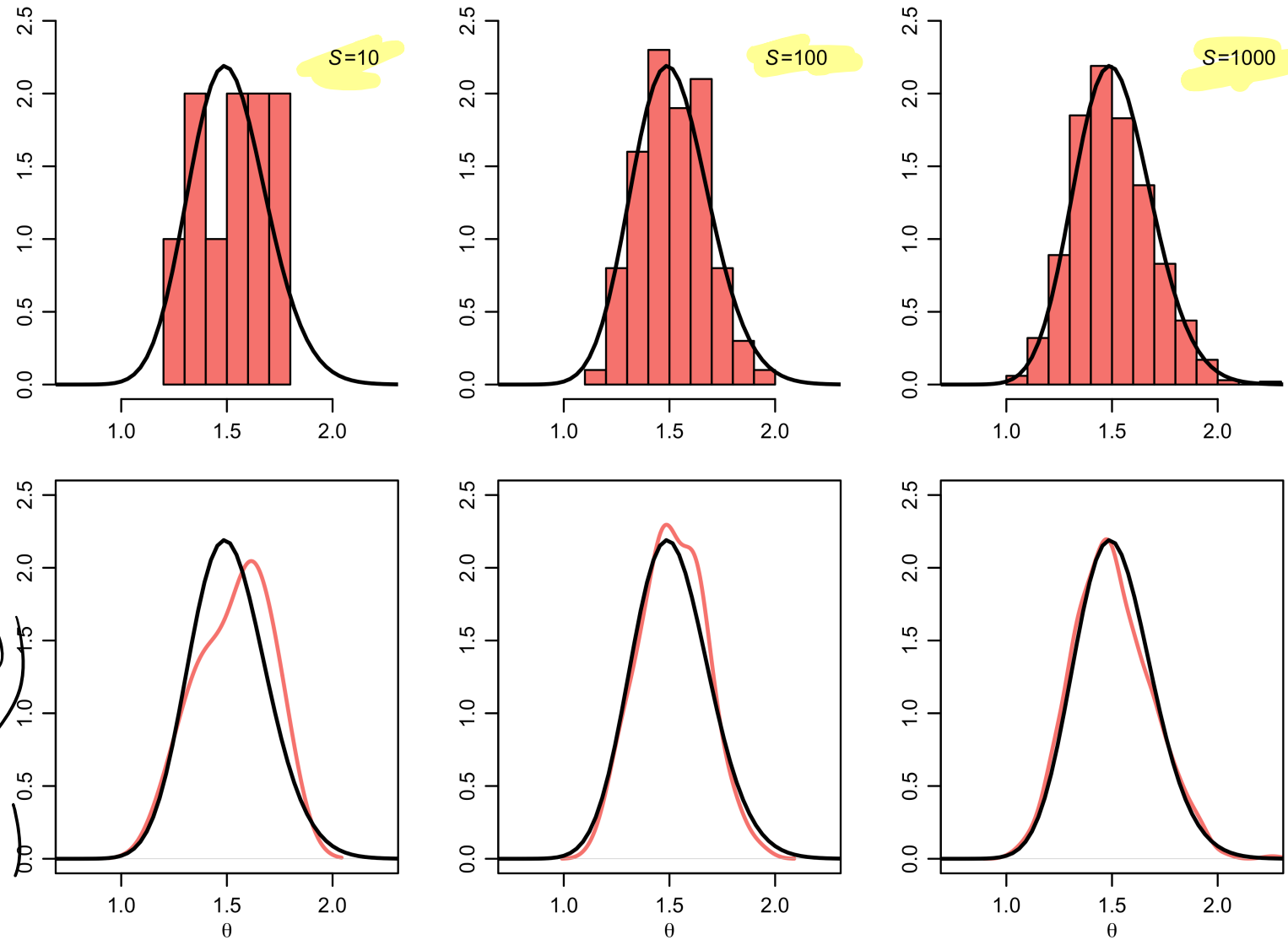
Professor Alexander Franks

Announcements

- Reading: Chapter 10 and 11

BDA

Monte Carlo approximations of a distribution



$$E[\theta \mid y] \approx \frac{1}{S}\sum \theta^{s}$$

$$Var\left(\frac{1}{S}\sum \theta^{(s)}\right) = \frac{V(\theta \mid y)}{S}$$

Decreasing M.C. Error →

Sampling strategies

$P(\theta_1, \theta_2 | y) = P(\theta_1 | y) P(\theta_2 | \theta_1, y)$

- Inversion Sampling (works for univariate distns)
- Grid sampling (works for low dimensional problems)
- Rejection sampling
- Importance sampling
- Markov Chain Monte Carlo

Monte Carlo to approximate integrals

$P(\theta | y) \propto L(\theta) P(\theta)$

$E[\theta | y]$? CI?

Sampling strategies

Why sampling?

Probability Integral Transform

- Suppose that a random variable, Y has a continuous distribution for with CDF is F_Y.
- Then the random variable $U = F_Y(Y)$ has a uniform distribution
  - This is known as the "probability integral transform PIT"
- By taking the inverse of F_Y we have $F^{-1}(U) = Y$

(Inversion Sampling)

Inversion Sampling

The inverse transform sampling method works as follows:

1. Generate a random number u from $\mathrm{Unif}[0, 1]$
2. Find the inverse of the desired CDF, e.g. $F_Y^{-1}(u)$.
3. Compute $y = F_Y^{-1}(u)$. y is now a sample from the desired distribution.

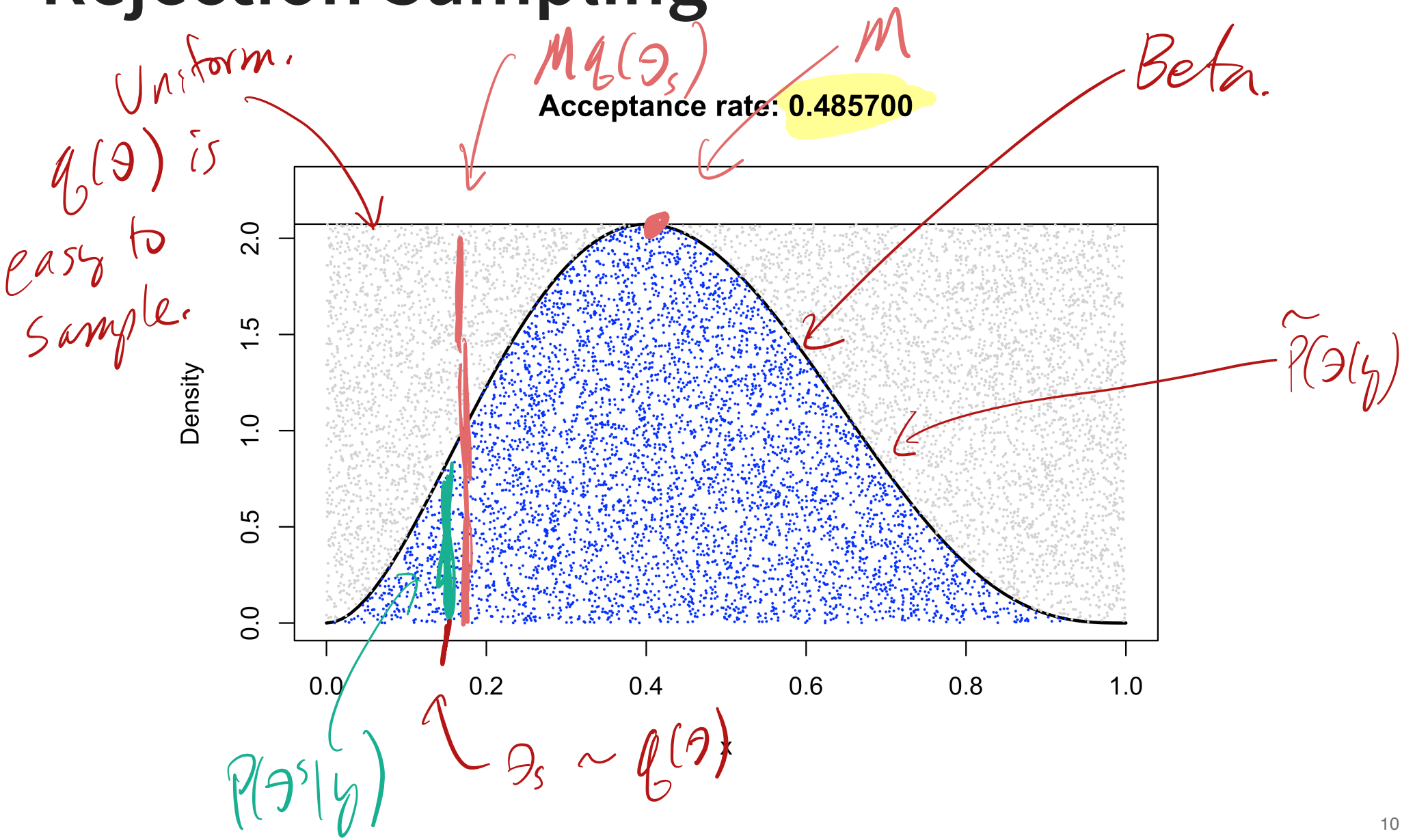
Inversion Sampling

Animation

Inversion Sampling

- Inversion sampling can be a fast and simple way to sample from a distribution
- Only effective if we know the inverse-CDF and can easily compute it
- This is a big challenge in practice. For example, even the normal distribution has a CDF, Φ, which cannot be expressed analytically.
  - Shifts from one hard problem (sampling) to another (computing an integral)
  - Need alternatives!

Rejection Sampling



Rejection Sampling algorithm

1. Choose a proposal density, $q(\theta)$ that we can easily sample from (e.g. uniform or normal) such that:

2. Find $M = \max \frac{p(\theta|y)}{q(\theta)}$

- If $M = \infty$ then q cannot be used as a proposal distribution
- If M is finite, $Mq(\theta)$ "envelopes" $p(\theta|y)$

3. Draw a sample, $\theta^{(s)}$ from $q(\theta)$

4. Accept $\theta^{(s)}$ as a draw from $p(\theta \mid y)$ with probability $\frac{p(\theta^{(s)}|y)}{Mq(\theta^{(s)})}$

Rejection Sampling

Demo

Rejection Sampling

- What is the expected number of rejections?
- How can we show this algorithm is valid?

Let Z be a random indicator for accepting θ^* from q targeting $P(\theta|y)$.

Expected # of acceptances:

$$P(Z=1) = E_q[Z] = E[E[Z|\theta^*]]$$

$$= E_q\left[\frac{P(\theta^*|y)}{M q(\theta^*)}\right] = \int \frac{P(\theta^*|y)}{M \cancel{q(\theta^*)}} \cancel{q(\theta^*)}\, d\theta^*$$

$$= \frac{1}{M}\int P(\theta^*|y)\, d\theta^* = \frac{1}{M}$$ how good is the match.

$$\tilde{P}(\theta|y) = L(\theta)P(\theta), \quad \text{Define } K = \int \tilde{P}(\theta|y)\, d\theta$$

$$\tilde{M} = \max \frac{\tilde{P}}{q} = \frac{P/K}{q}, \quad \tilde{M} = M/K$$

$$E_q\left[\frac{\tilde{P}}{\tilde{M} q}\right] = \frac{K}{\tilde{M}} = \frac{1}{M}$$

$$\begin{cases} \theta^* \sim q, \text{ consider } (Z, \theta^*) \\ Z|\theta^* \sim \text{Bern}\left(\frac{\tilde{P}(\theta|y)}{\tilde{M} q(\theta)}\right) \end{cases}$$

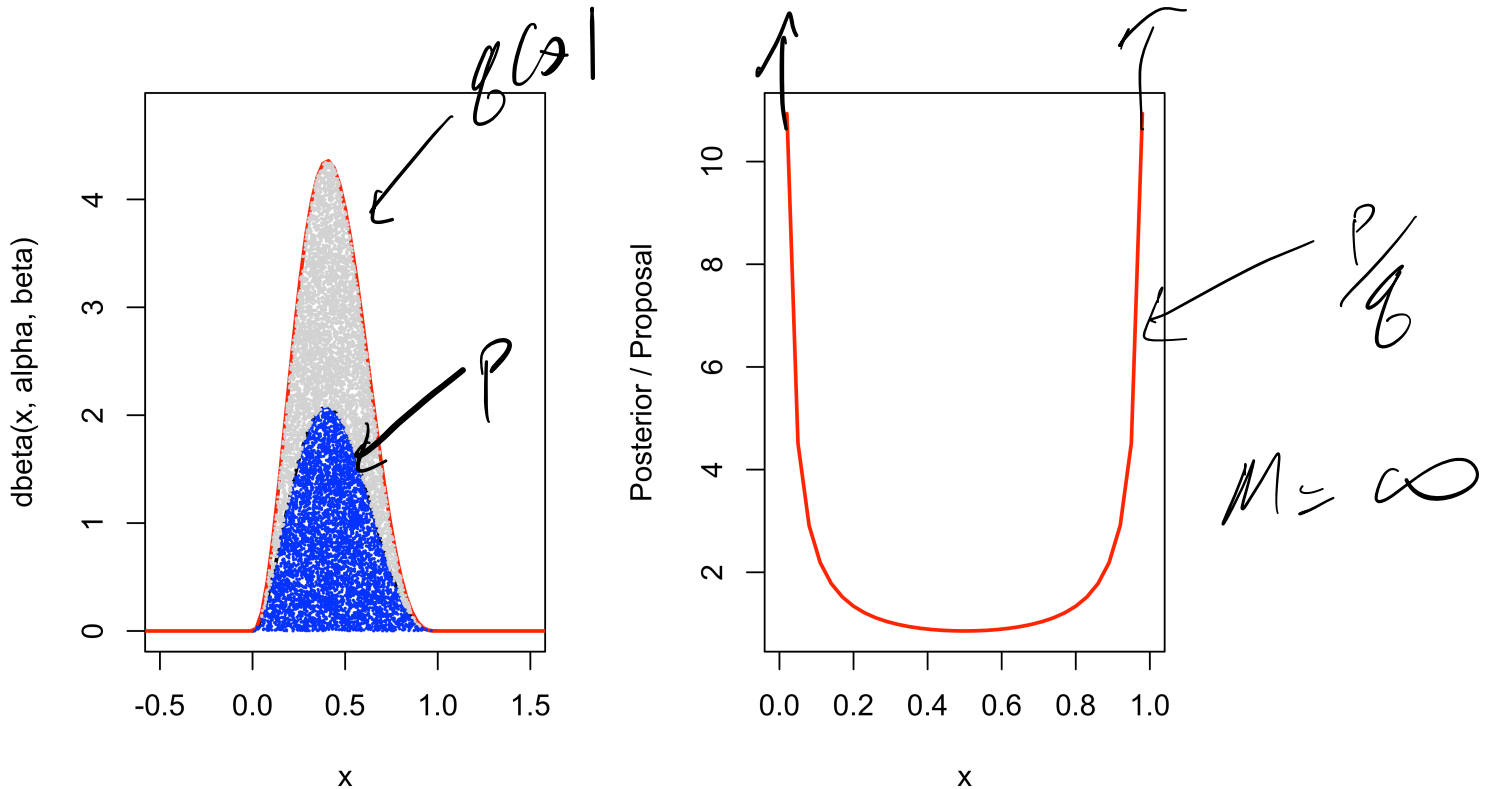
$$P(Z, \theta^*) = q(\theta)\left(\frac{\tilde{P}(\theta|y)}{\tilde{M} q(\theta)}\right)^Z \left(1 - \frac{\tilde{P}(\theta|y)}{\tilde{M} q(\theta)}\right)^{1-Z}$$

$$P(\theta|Z=1) = \frac{P(Z=1, \theta)}{P(Z=1)} = \frac{\cancel{q(\theta)} \frac{\tilde{P}(\theta|y)}{\cancel{\tilde{M} q(\theta)}}}{K/\cancel{\tilde{M}}}$$

$$= \frac{\tilde{p}(\theta|y)}{K} = p(\theta|y)$$

Proposal most envelope target

Sometimes its not obvious…



Target density in black, proposal density in red

The German Tank Problem

The German Tank Problem

- US and allies wanted to estimate how many tanks were produced by the Germans during WWII, n?
- They used the serial numbers on captured or destroyed tanks to estimate tank production
- Assuming serial numbers, Z_i, are ordered from 1 to n, where n is the number of tanks produced
- How do we estimate n?

The German Tank Problem

- Assume we observe the serial number of only one tank, y.
- What is a plausible sampling model? How do we write the likelihood?
- What is a reasonable prior?

Observe $y \sim \text{Discrete Unif } 1, \ldots, n.$

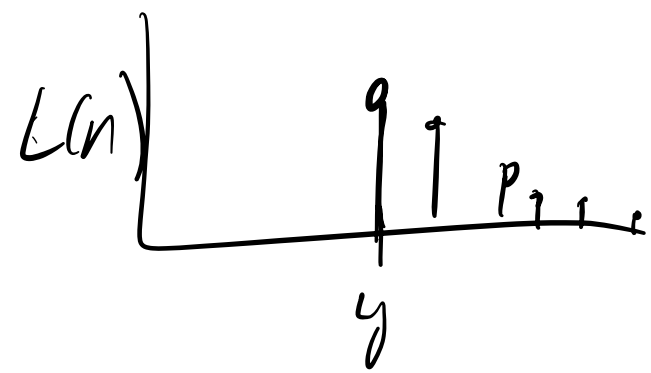
$P(n)$?

The German Tank Problem

- Assume y is uniformly distributed between 1 and n
- Specify a Poisson prior distribution
  - Assume a priori a rate of 200 tanks / month
- Use Monte Carlo (rejection sampler) to summarize the results. What is a reasonable proposal?

MLE for n?

$$L(n) = I[n \geq y]\frac{1}{n}$$



$$\hat{n}_{MLE} = y.$$

$$P(n|y) \propto \frac{1}{n} I[n \geq y] \frac{e^{-\lambda} \lambda^n}{n!}$$

Propose from $q \sim \text{Pois}(\lambda)$

$$\tilde{m} = \max_n \frac{P(n|y)}{q(n)} = \max_n \frac{\frac{1}{n} \frac{e^{-\lambda n}}{n!} I[n \geq y]}{\frac{e^{-\lambda} \lambda^n}{n!}}$$

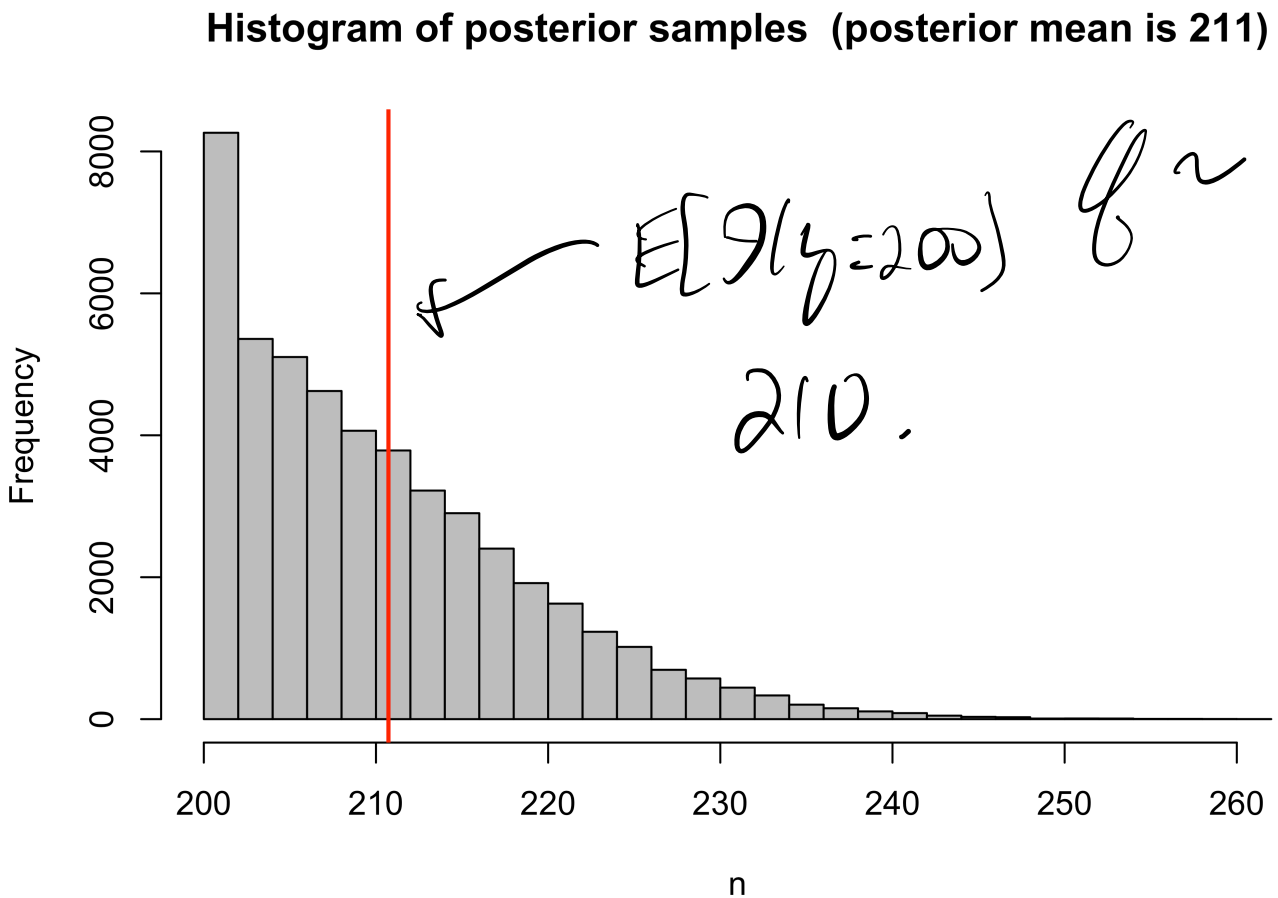
$$= \max_n \frac{1}{n} I[n \geq y]$$

$$= 1/y.$$

Propose $q \sim \text{Pois}(\lambda)$

Acc. w.p. $\frac{\tilde{P}(n|y)}{\tilde{m} q} = \frac{y}{n} I[n \geq y]$

The German Tank Problem, $y = 200$



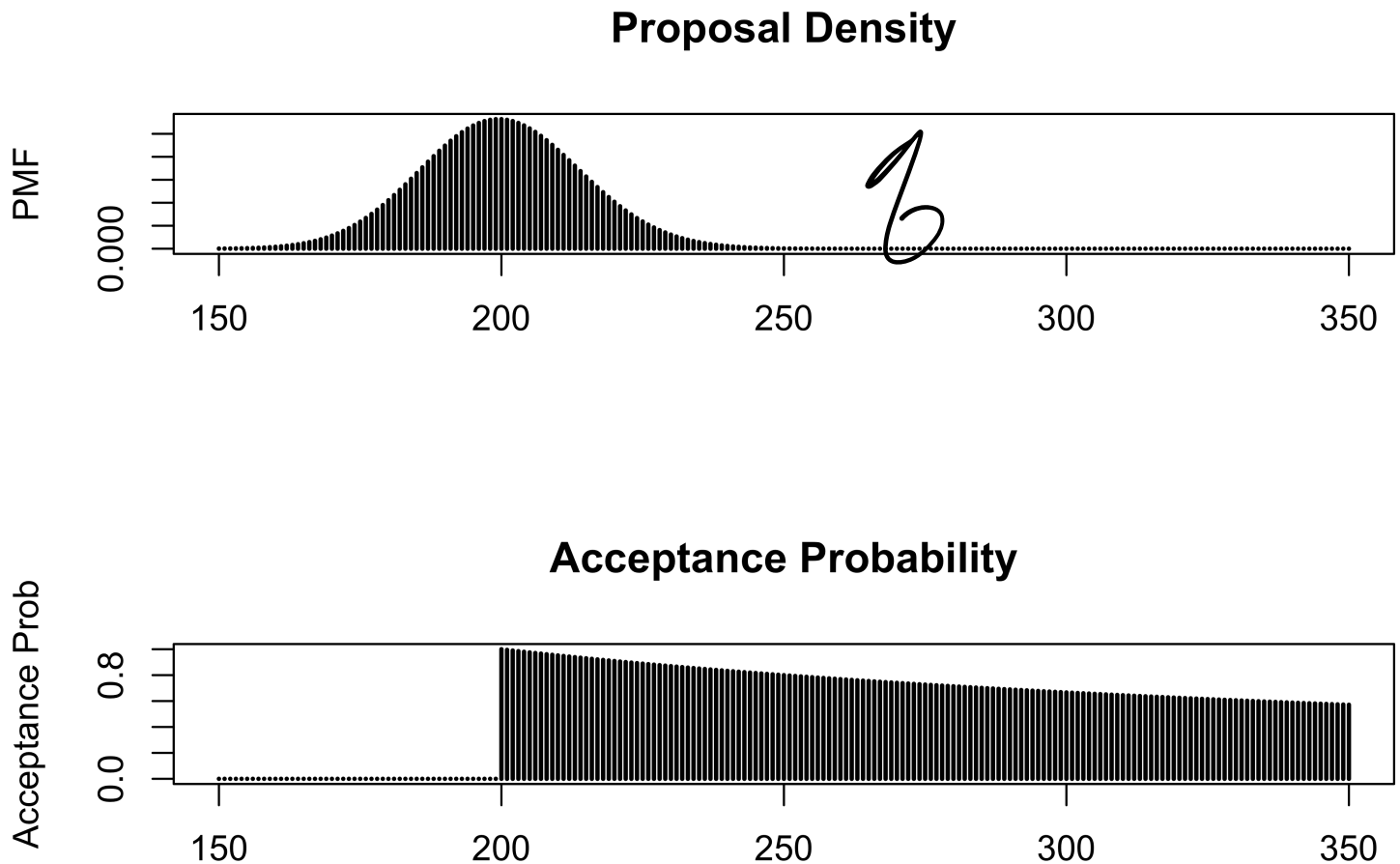
$E[\theta \mid y=200]$

210.

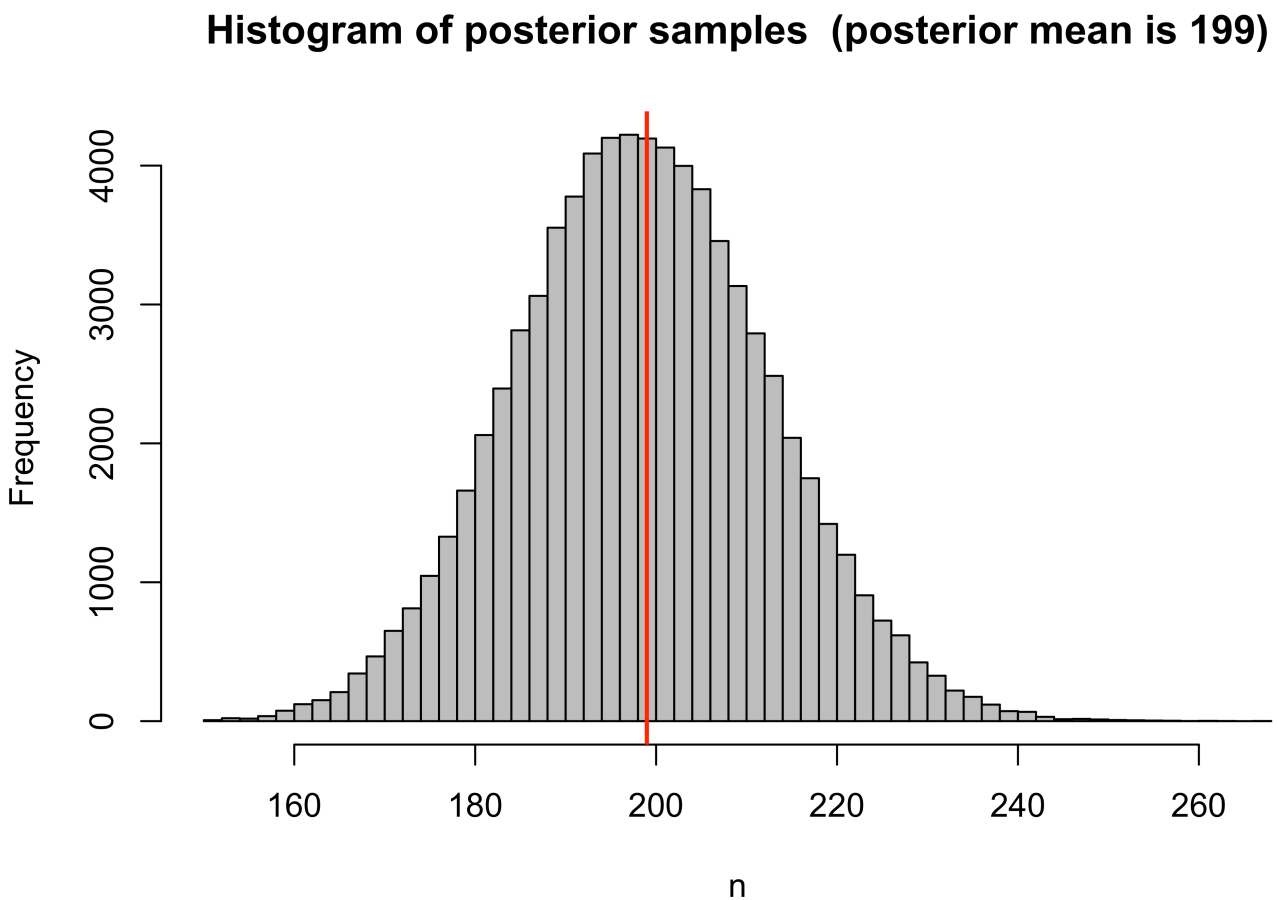
$\theta \sim \text{Pois}(200)$

Acceptance rate: 0.48257

The German Tank Problem, Z = 200

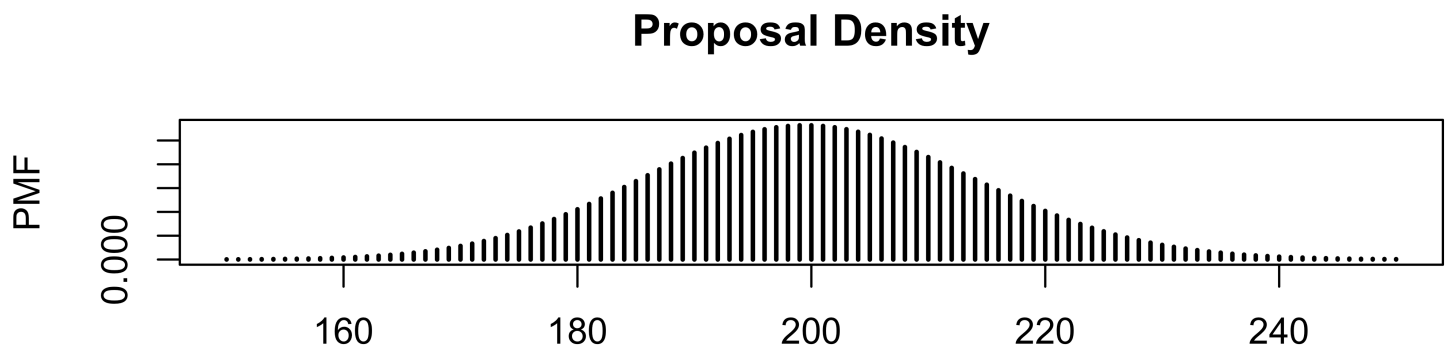


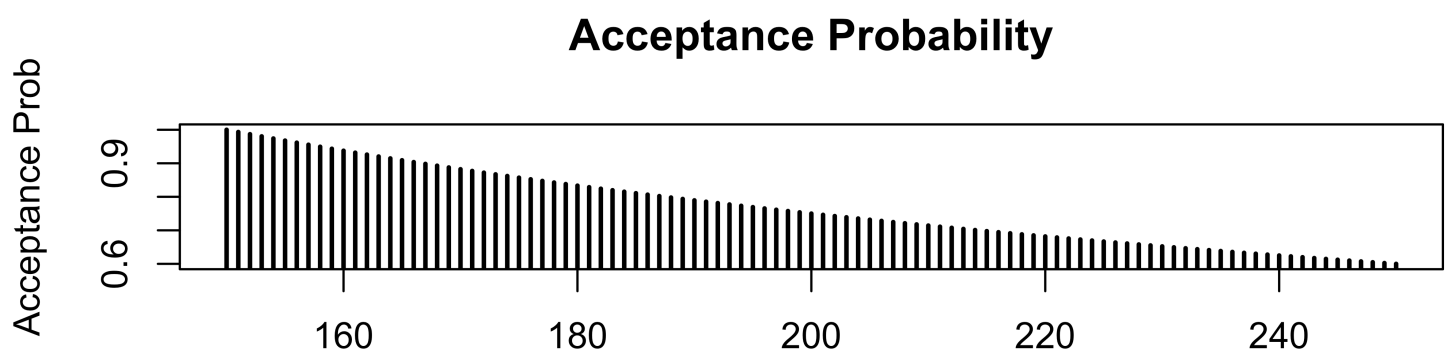
The German Tank Problem, Z = 150



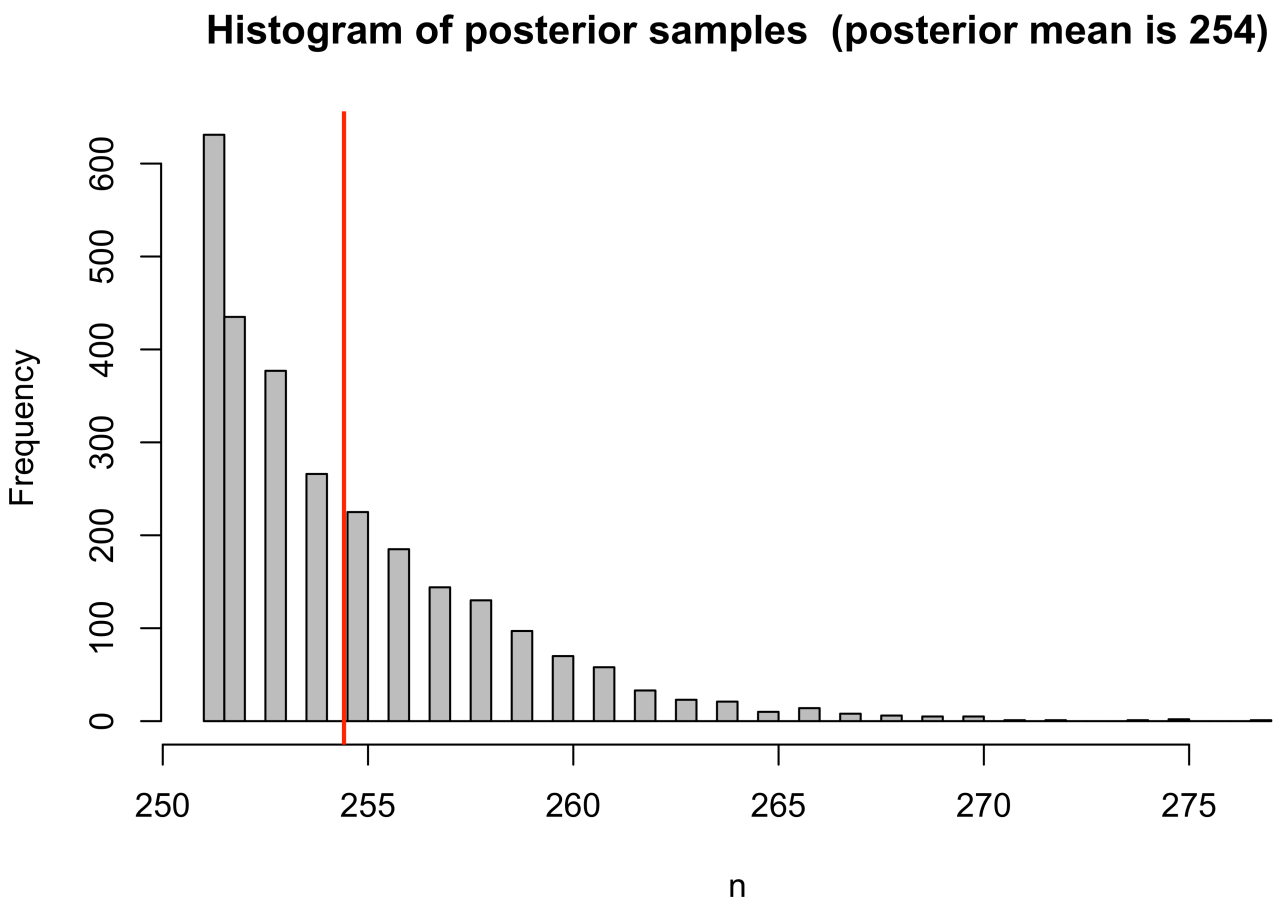
Acceptance rate: 0.75292

German Tank Problem, Z=150



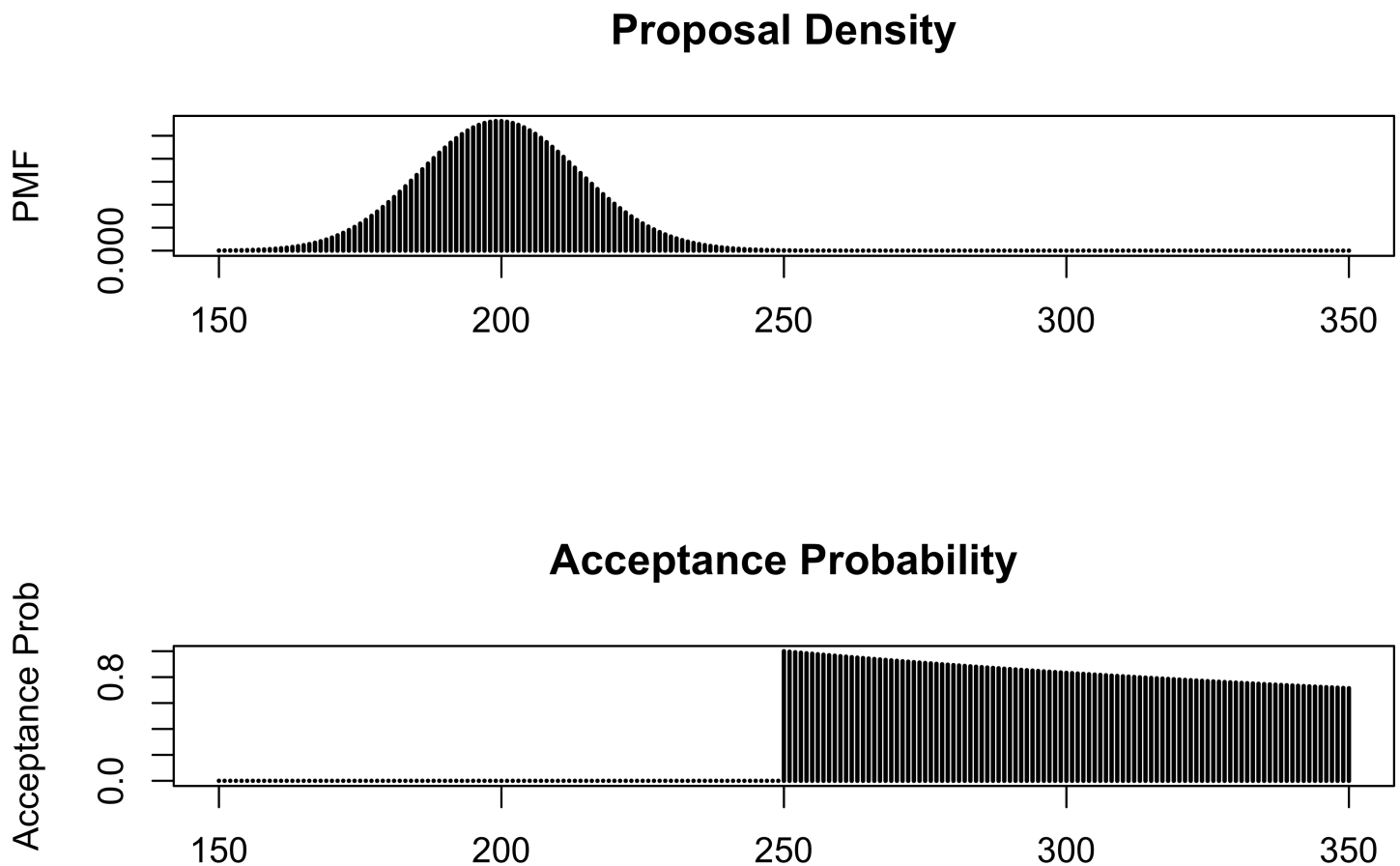


The German Tank Problem, Z = 250



Acceptance rate: 2.749^{-4}

The German Tank Problem, Z = 250



Importance Sampling

- With rejection sampling we get actual samples from the target density
- Downside: often throw away many samples. Inefficient if rejection rate is high.
  - Main challenge is finding a valid proposal density and one that minimzes the number of rejections
- Importance sampling: use all samples from proposal to compute expected values.
- Weight the *importance* of each sample by how representative it is of the target distribution

Want $E_{P(\theta|y)}[h(\theta)] \overset{\text{LOTUS}}{=}$

Too Hard.

$\theta^* \sim P(\theta|y)$

$\frac{1}{S}\sum h(\theta^*)$

$$\int_{\Theta} h(\theta)\, P(\theta|y)\, d\theta =$$

$$\int h(\theta)\, P(\theta|y)\, \frac{q(\theta)}{q(\theta)}\, d\theta = \int h(\theta)\, \underbrace{\frac{P(\theta|y)}{q(\theta)}}_{w(\theta)}\, q(\theta)\, d\theta$$

$$E_q[w(\theta)h(\theta)] \approx \boxed{\frac{1}{S}\sum w(\theta_s)\, h(\theta_s) \quad \text{where } \theta_s \sim q(\theta)}$$

$$E_q[w(\theta)] = \int \frac{P(\theta|y)}{\cancel{q(\theta)}}\, \cancel{q(\theta)} = 1$$

If $\tilde{P}$, $\longrightarrow E[w] = K$

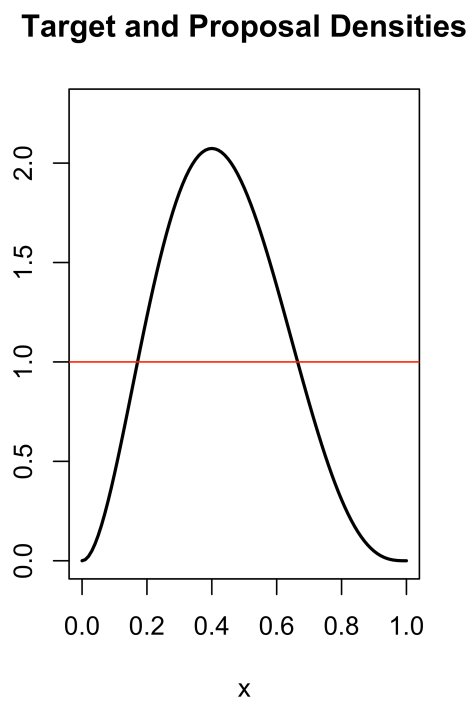
In practice $w(\theta) = \dfrac{\tilde{P}(\theta|y)}{q(\theta)} \Big/ \sum_s \dfrac{\tilde{P}(\theta|y)}{q(\theta)} \approx K$

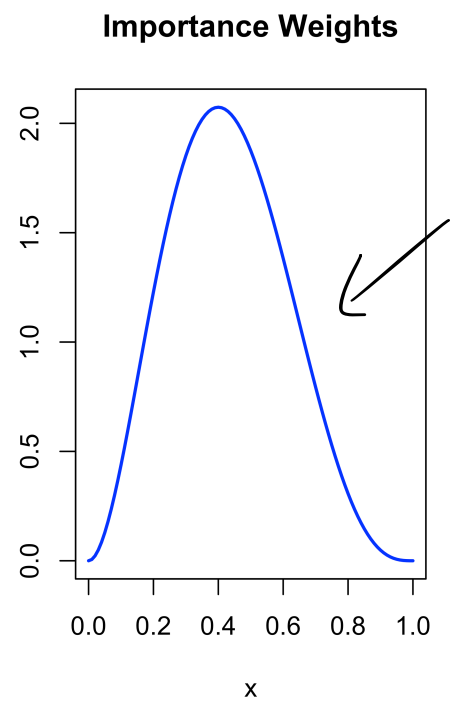
Importance Sampling

- Examine histogram of log importance weights
- Want the largest weights to be not too large
- Effective sample size: $\frac{1}{\sum_{s=1}^{S} \tilde{w}(\theta^s)^2}$ where $\tilde{w}(\theta^s)$ are the normalized weights

Importance Sampling

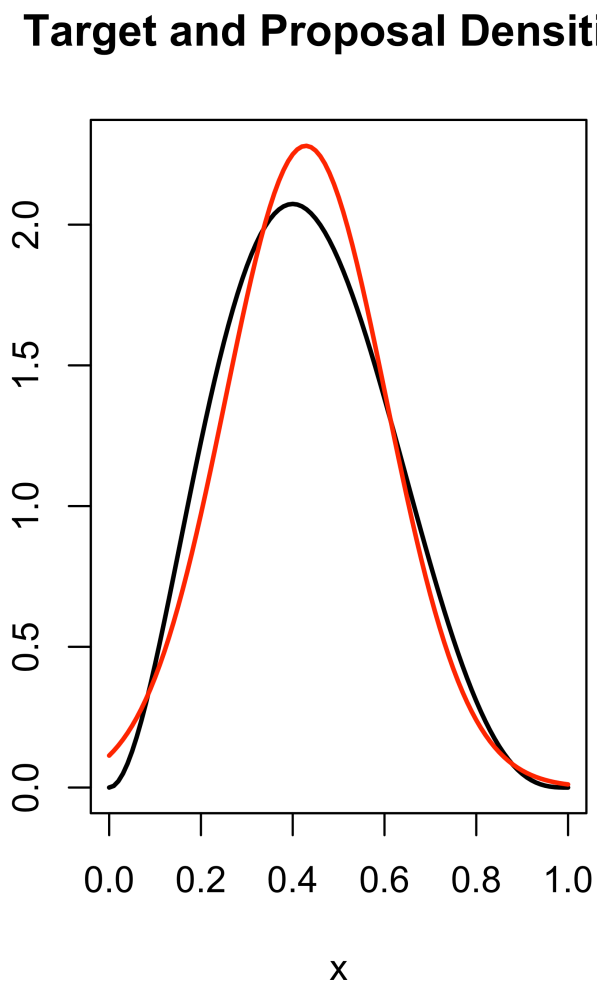
- Target: a Beta(3, 4)
- Proposal: Uniform

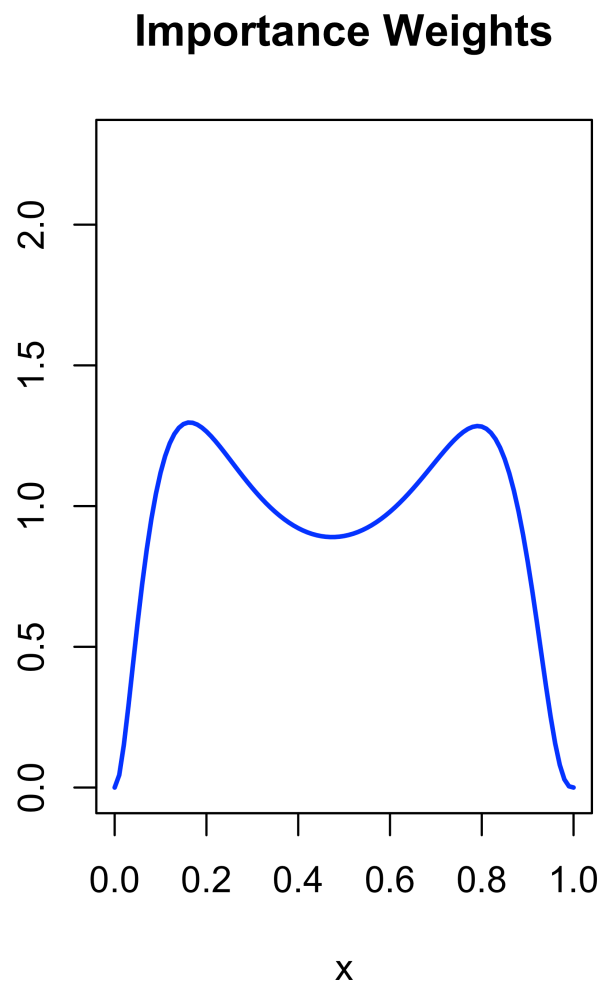




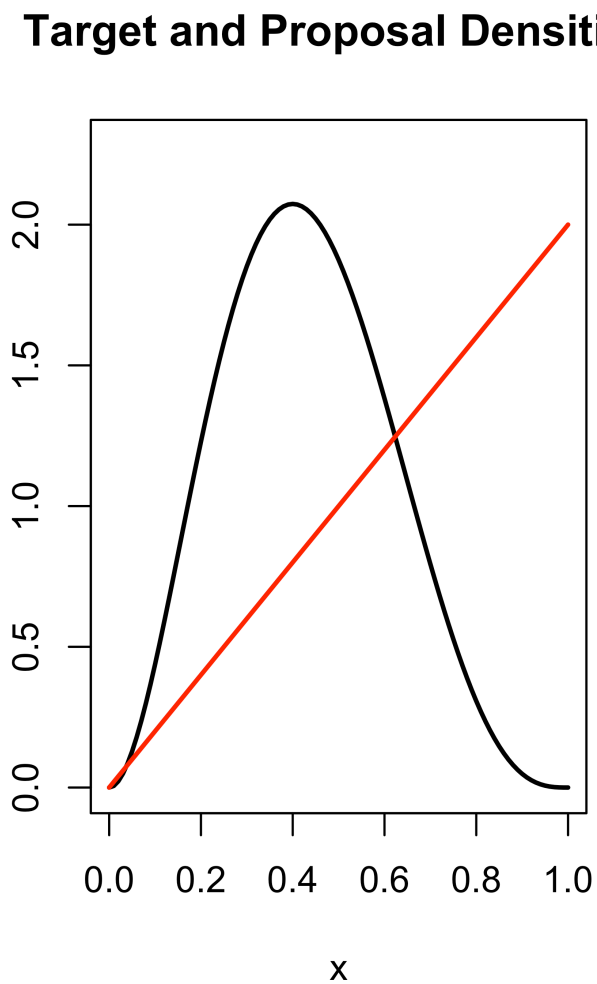
$\frac{p}{q}$

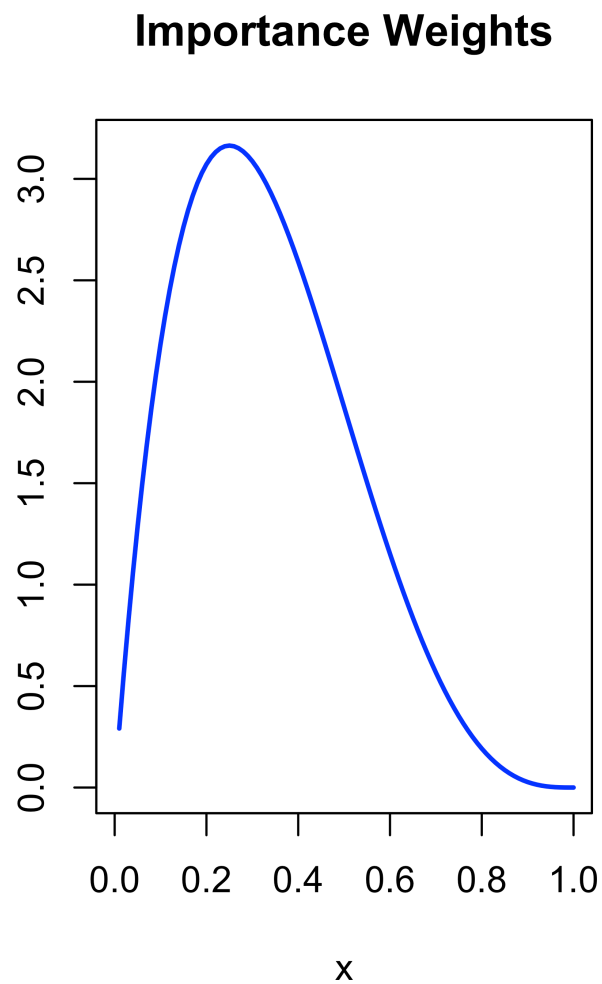
Importance Sampling





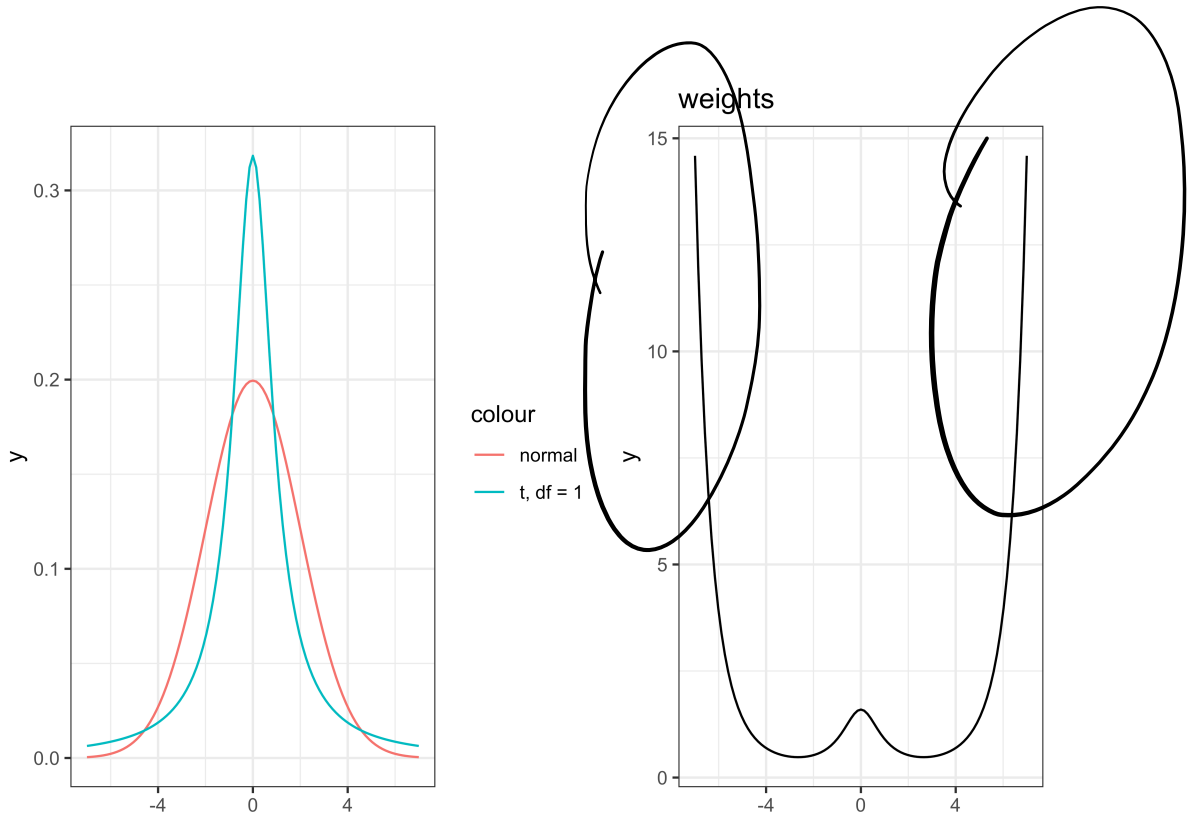
Importance Sampling





Importance Sampling

- Target: Cauchy(0, 1)
- Proposal: Normal(0, 3)



Importance Sampling

```r
df <- 1

proposal_draws <- rnorm(1e5, sd=2)
proposal_draws <- sort(proposal_draws)

w <- dt(proposal_draws, df=df) / dnorm(proposal_draws, sd=2)
w <- w/sum(w)

q <- 0.25
idx <- match(FALSE, cumsum(w) < q)
qt(q, df=df)
```

Cauchy.

```
[1] -1
```

```r
proposal_draws[idx]
```

```
[1] -0.860783
```

```r
q <- 0.05
idx <- match(FALSE, cumsum(w) < q)
qt(q, df=df)
```

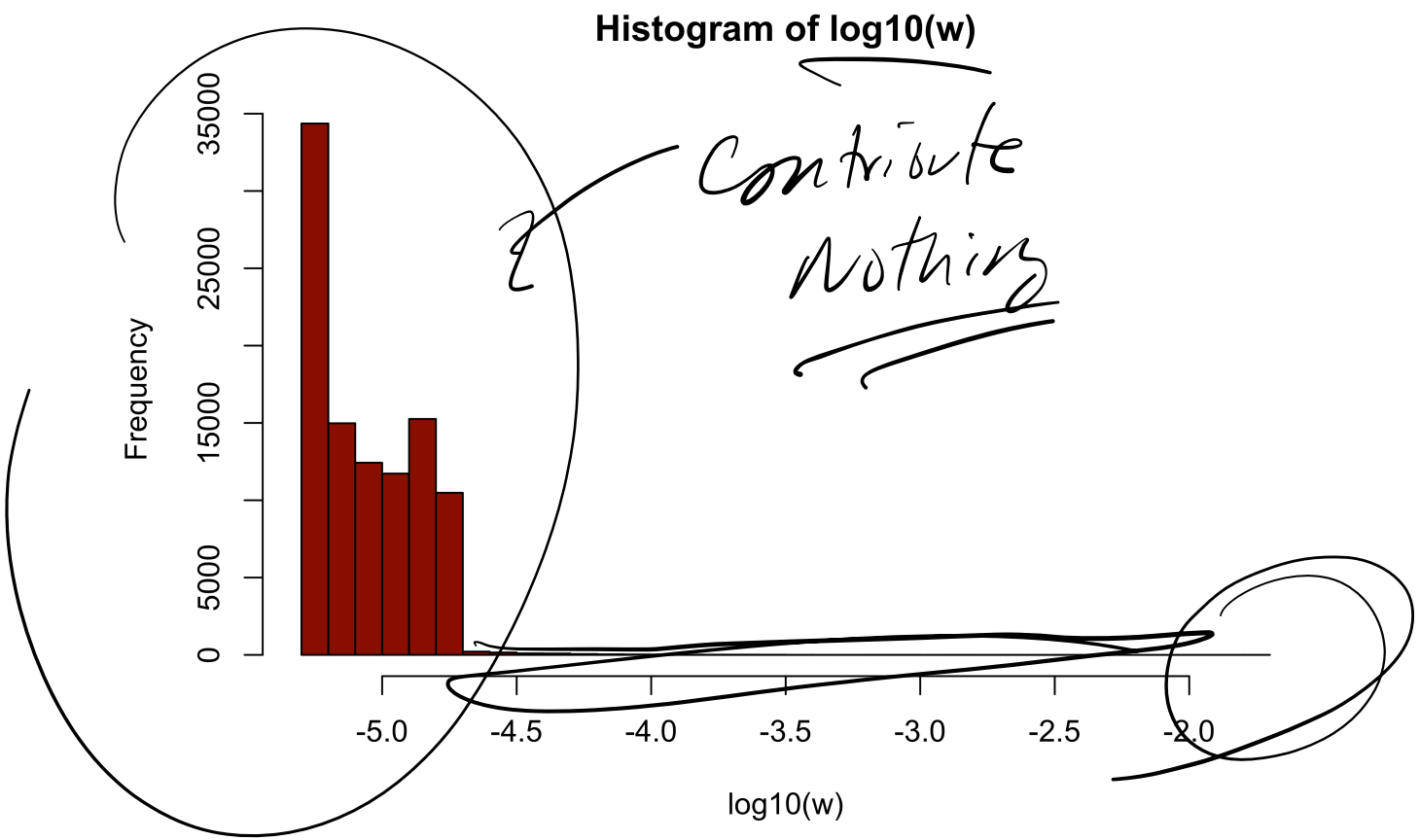
```
[1] -6.313752
```

```r
proposal_draws[idx]
```

Importance Sampling

```
[1] "99% quantile of log10(w) is -4.773680"
[1] "max of log10(w) is -1.769665"
```

Look for Skew.



Contribute Nothing

A Gamma Puzzle

- Goal: report the exact mean of a Gamma(a, b) using only one random draw
- Allowed to generate a random draw from a proposal distribution of your choosing
- You can evaluate the gamma density and the proposal density at the generated value

Want: $\int_0^\infty x\,p(x)\,dx = \int_0^\infty x \frac{b^a e^{-bx} x^{a-1}}{\Gamma(a)}\,dx$

$q(x) \sim \text{Gam}(a+1, b)$

$$= \int_0^\infty \underbrace{\frac{\cancel{x}\, b^a \cancel{e^{-bx}} \cancel{x^{a-1}}}{\Gamma(a)} \Big/ \frac{b^{a+1} \cancel{e^{-bx}} \cancel{x^a}}{\Gamma(a+1)}}_{h(x)\,L(x)} \underbrace{\frac{b^{a+1} e^{-bx} x^a}{\Gamma(a+1)}}_{q(x)}$$

A Gamma Puzzle

```r
## Generate 10 draws from a gamma 5, 2
proposal_draws <- rgamma(10, shape = 5, rate=2)

## Targt a gamma 4, 2
w <-
  dgamma(proposal_draws, shape=4, rate=2) /
  dgamma(proposal_draws, shape=5, rate=2)

w*proposal_draws
```

```
[1] 2 2 2 2 2 2 2 2 2 2
```

Want $Var(w(\theta))$ small.

Really want $Var(h(\theta)w(\theta))$ small.