# ECON5120 Bayesian Data Analysis

## Week 2: Normal model and Posterior Expectations

Santiago Montoya-Blandón

University of Glasgow
Adam Smith Business School

January 21, 2026

# Outline

# Normal model

- Likelihood: $y_i \sim \mathcal{N}(\mu, \sigma^2)$ for $i = 1, \ldots, n$. Let $y = (y_1, \ldots, y_n)$:

$$p(y|\mu, \sigma^2) = \prod_{i=1}^{n} p(y_i|\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\right\}$$

- Non-informative prior: $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$
- *Joint* posterior: $p(\mu, \sigma^2|y) \propto p(y|\mu, \sigma^2)p(\mu, \sigma^2)$

# Normal model

- Likelihood: $y_i \sim \mathcal{N}(\mu, \sigma^2)$ for $i = 1, \ldots, n$. Let $y = (y_1, \ldots, y_n)$:

$$p(y|\mu, \sigma^2) = \prod_{i=1}^{n} p(y_i|\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\right\}$$

- Non-informative prior: $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$
- *Joint* posterior: $p(\mu, \sigma^2|y) \propto p(y|\mu, \sigma^2)p(\mu, \sigma^2)$
- For simplicity in what follows, define

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\,, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \quad \text{and} \quad \bar{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

# Normal model: Details

$$p(\mu, \sigma^2|y) \propto \sigma^{-2} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\right\}$$

# Normal model: Details

$$p(\mu, \sigma^2|y) \propto \sigma^{-2} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\right\}$$

$$\propto \sigma^{-n-2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right\}$$

# Normal model: Details

$$
\begin{aligned}
p(\mu, \sigma^2|y) &\propto \sigma^{-2} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\right\} \\
&\propto \sigma^{-n-2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right\} \\
&= \sigma^{-n-2} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right]\right\}
\end{aligned}
$$

# Normal model: Details

$$
\begin{aligned}
p(\mu, \sigma^2|y) &\propto \sigma^{-2} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \mu)^2\right\} \\
&\propto \sigma^{-n-2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right\} \\
&= \sigma^{-n-2} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right]\right\} \\
&= \sigma^{-n-2} \exp\left\{-\frac{1}{2\sigma^2}\left[n\bar{s}^2 + n(\bar{y} - \mu)^2\right]\right\}
\end{aligned}
$$

# Normal model: Details

$$
\begin{aligned}
\sum_{i=1}^{n}(y_i-\mu)^2 &= \sum_{i=1}^{n}(y_i^2-2y_i\mu+\mu^2)\\
&= \sum_{i=1}^{n}(y_i^2-2y_i\mu+\mu^2-\bar{y}^2+\bar{y}^2-2y_i\bar{y}+2y_i\bar{y})\\
&= \sum_{i=1}^{n}(y_i^2-2y_i\bar{y}+\bar{y}^2)+\sum_{i=1}^{n}(\mu^2-2y_i\mu-\bar{y}^2+2y_i\bar{y})\\
&= \sum_{i=1}^{n}(y_i-\bar{y})^2+n(\mu^2-2\bar{y}\mu-\bar{y}^2+2\bar{y}\bar{y})\\
&= \sum_{i=1}^{n}(y_i-\bar{y})^2+n(\bar{y}-\mu)^2
\end{aligned}
$$

# Normal model: Joint posterior

- Rewrite the final expression to obtain

$$p(\mu, \sigma^2|y) \propto \underbrace{\exp\left\{-\frac{(\mu-\bar{y})^2}{2(\sigma^2/n)}\right\}}_{p(\mu|\sigma^2,y)} \times \underbrace{\left(\sigma^2\right)^{-n/2-1} \exp\left\{-\frac{n\bar{s}^2}{2\sigma^2}\right\}}_{p(\sigma^2|y)}$$

# Normal model: Joint posterior

- Rewrite the final expression to obtain

$$p(\mu, \sigma^2|y) \propto \underbrace{\exp\left\{-\frac{(\mu-\bar{y})^2}{2(\sigma^2/n)}\right\}}_{p(\mu|\sigma^2,y)} \times \underbrace{\left(\sigma^2\right)^{-n/2-1} \exp\left\{-\frac{n\bar{s}^2}{2\sigma^2}\right\}}_{p(\sigma^2|y)}$$

- These are the kernels for two known distributions
- We have $p(\mu, \sigma^2|y) = p(\mu|\sigma^2, y)p(\sigma^2|y)$ with

$$p(\mu|\sigma^2, y) = \mathcal{N}\left(\mu \,\middle|\, \bar{y}, \frac{\sigma^2}{n}\right)$$

$$p(\sigma^2|y) = \text{Inv-Gamma}\left(\sigma^2 \,\middle|\, \frac{n}{2}, \frac{n\bar{s}^2}{2}\right)$$

# Normal model: Marginal posteriors

- We can also learn the *marginal* posteriors $p(\mu|y)$ and $p(\sigma^2|y)$:

$$p(\mu|y) = \int p(\mu, \sigma^2|y) d\sigma^2 \quad \text{and} \quad p(\sigma^2|y) = \int p(\mu, \sigma^2|y) d\mu$$

- These can be found to be as follows

$$\mu|y \sim t_{n-1}\left(\bar{y}, \frac{s^2}{n}\right)$$

$$\sigma^2|y \sim \text{Inv-Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$$

# Normal model: Marginal posteriors

- We can also learn the *marginal* posteriors $p(\mu|y)$ and $p(\sigma^2|y)$:

$$p(\mu|y) = \int p(\mu, \sigma^2|y) d\sigma^2 \quad \text{and} \quad p(\sigma^2|y) = \int p(\mu, \sigma^2|y) d\mu$$

- These can be found to be as follows

$$\mu|y \sim t_{n-1}\left(\bar{y}, \frac{s^2}{n}\right)$$

$$\sigma^2|y \sim \text{Inv-Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$$

- Note that marginalizing introduces extra uncertainty through thicker tails:
  1. For $\mu$, the distribution is still centered at $\bar{y}$ but it's now a $t$ with $n-1$ degrees of freedom instead of a normal
  2. For $\sigma^2$, the Inv-Gamma distribution has one less degree of freedom

# Normal model: Marginal posterior for $\sigma^2$

$$p(\sigma^2|y) = \int p(\mu, \sigma^2|y) d\mu$$

# Normal model: Marginal posterior for $\sigma^2$

$$p(\sigma^2|y) = \int p(\mu, \sigma^2|y) d\mu$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{n\bar{s}^2}{2\sigma^2}\right\} \int \exp\left\{-\frac{(\mu-\bar{y})^2}{2(\sigma^2/n)}\right\} d\mu$$

# Normal model: Marginal posterior for $\sigma^2$

$$
\begin{aligned}
p(\sigma^2|y) &= \int p(\mu, \sigma^2|y) d\mu \\
&\propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{n\bar{s}^2}{2\sigma^2}\right\} \int \exp\left\{-\frac{(\mu-\bar{y})^2}{2(\sigma^2/n)}\right\} d\mu \\
&= \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{n\bar{s}^2}{2\sigma^2}\right\} \sqrt{2\pi\left(\frac{\sigma^2}{n}\right)} \int \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{n}\right)}} \exp\left\{-\frac{(\mu-\bar{y})^2}{2(\sigma^2/n)}\right\} d\mu
\end{aligned}
$$

# Normal model: Marginal posterior for $\sigma^2$

$$
\begin{aligned}
p(\sigma^2|y) &= \int p(\mu, \sigma^2|y) d\mu \\
&\propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{n\bar{s}^2}{2\sigma^2}\right\} \int \exp\left\{-\frac{(\mu-\bar{y})^2}{2(\sigma^2/n)}\right\} d\mu \\
&= \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{n\bar{s}^2}{2\sigma^2}\right\} \sqrt{2\pi\left(\frac{\sigma^2}{n}\right)} \int \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{n}\right)}} \exp\left\{-\frac{(\mu-\bar{y})^2}{2(\sigma^2/n)}\right\} d\mu \\
&\propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} (\sigma^2)^{1/2} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\}
\end{aligned}
$$

# Normal model: Marginal posterior for $\sigma^2$

$$
\begin{aligned}
p(\sigma^2|y) &= \int p(\mu, \sigma^2|y) d\mu \\
&\propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{n\bar{s}^2}{2\sigma^2}\right\} \int \exp\left\{-\frac{(\mu-\bar{y})^2}{2(\sigma^2/n)}\right\} d\mu \\
&= \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{n\bar{s}^2}{2\sigma^2}\right\} \sqrt{2\pi\left(\frac{\sigma^2}{n}\right)} \int \frac{1}{\sqrt{2\pi\left(\frac{\sigma^2}{n}\right)}} \exp\left\{-\frac{(\mu-\bar{y})^2}{2(\sigma^2/n)}\right\} d\mu \\
&\propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} (\sigma^2)^{1/2} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\} \\
&= \left(\frac{1}{\sigma^2}\right)^{(n-1)/2+1} \exp\left\{-\frac{(n-1)s^2}{2\sigma^2}\right\}
\end{aligned}
$$

# Normal model: Marginal posterior for $\mu$

$$p(\mu|y) = \int p(\mu, \sigma^2|y) d\sigma^2$$

# Normal model: Marginal posterior for $\mu$

$$
\begin{aligned}
p(\mu|y) &= \int p(\mu, \sigma^2|y) d\sigma^2 \\
&\propto \int \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\mu - \bar{y})^2\right]\right\} d\sigma^2
\end{aligned}
$$

# Normal model: Marginal posterior for $\mu$

$$
\begin{aligned}
p(\mu|y) &= \int p(\mu, \sigma^2|y) d\sigma^2 \\
&\propto \int \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\mu-\bar{y})^2\right]\right\} d\sigma^2 \\
&\propto \left[(n-1)s^2 + n(\mu-\bar{y})^2\right]^{-n/2}
\end{aligned}
$$

# Normal model: Marginal posterior for $\mu$

$$
\begin{aligned}
p(\mu|y) &= \int p(\mu, \sigma^2|y) d\sigma^2 \\
&\propto \int \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\mu-\bar{y})^2\right]\right\} d\sigma^2 \\
&\propto \left[(n-1)s^2 + n(\mu-\bar{y})^2\right]^{-n/2} \\
&= \left[(n-1)s^2\right]^{-n/2} \left[1 + \frac{n(\mu-\bar{y})^2}{(n-1)s^2}\right]^{-n/2}
\end{aligned}
$$

# Normal model: Marginal posterior for $\mu$

$$
\begin{aligned}
p(\mu|y) &= \int p(\mu, \sigma^2|y) d\sigma^2 \\
&\propto \int \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\mu-\bar{y})^2\right]\right\} d\sigma^2 \\
&\propto \left[(n-1)s^2 + n(\mu-\bar{y})^2\right]^{-n/2} \\
&= \left[(n-1)s^2\right]^{-n/2} \left[1 + \frac{n(\mu-\bar{y})^2}{(n-1)s^2}\right]^{-n/2} \\
&\propto \left[1 + \frac{1}{(n-1)} \frac{(\mu-\bar{y})^2}{(s^2/n)}\right]^{-[(n-1)+1]/2}
\end{aligned}
$$

# Normal model: Details of marginal posterior for $\mu$

- Define
$$A = (n-1)s^2 + n(\mu - \bar{y})^2 \quad \text{and} \quad u = \frac{A}{2\sigma^2}$$
- Substituting in the integral in the second line, we can express it as
$$\left(\frac{2}{A}\right)^{n/2} \int_0^\infty u^{n/2-1} \exp(-u) du$$
- Using the definition of a gamma integral $\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x) dx$, we can finally write
$$\left(\frac{2}{A}\right)^{n/2} \Gamma\left(\frac{n}{2}\right) \propto A^{-n/2}$$

# Normal model: Conjugate prior

- So far, we have used the non-informative prior $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$
- Conjugate prior has to have a form $p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$ (BDA ch. 3)

# Normal model: Conjugate prior

- So far, we have used the non-informative prior $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$
- Conjugate prior has to have a form $p(\mu, \sigma^2) = p(\mu|\sigma^2)p(\sigma^2)$ (BDA ch. 3)
- Handy parameterization

$$\mu|\sigma^2 \sim \mathcal{N}\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right)$$

$$\sigma^2 \sim \text{Inv-Gamma}\left(\frac{\alpha_0}{2}, \frac{\delta_0}{2}\right)$$

- $\mu$ and $\sigma^2$ are a priori dependent: if $\sigma^2$ is large, then $\mu$ has a wide prior

# Normal model: Conjugate prior

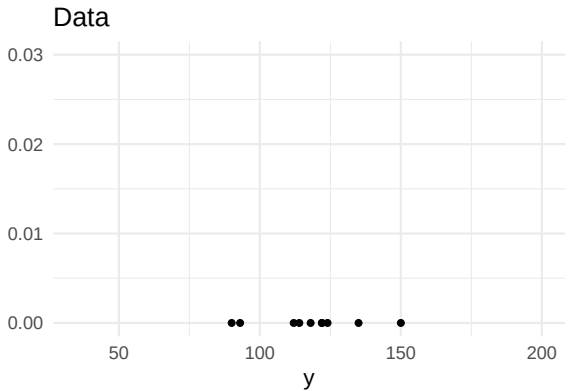
- The joint posterior has the same form as with the non-informative prior
- We can show (great exercise to do so!) that this joint posterior is:

$$\begin{aligned} p(\mu, \sigma^2|y) &= p(\mu|\sigma^2, y)p(\sigma^2|y) \\ &= \mathcal{N}\left(\mu \middle| \mu_n, \frac{\sigma^2}{\kappa_n}\right) \text{Inv-Gamma}\left(\sigma^2 \middle| \frac{\alpha_n}{2}, \frac{\delta_n}{2}\right) \end{aligned}$$

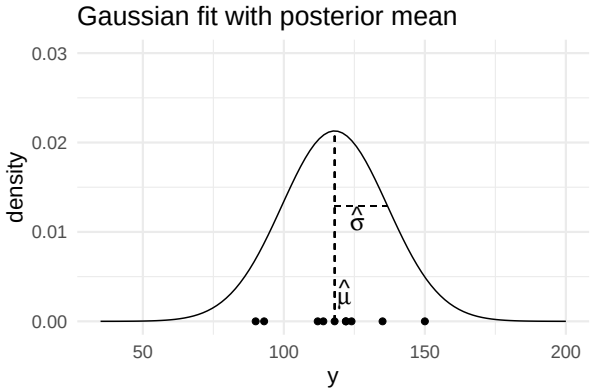
where

$$\begin{aligned} \mu_n &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y} \\ \kappa_n &= \kappa_0 + n \\ \alpha_n &= \alpha_0 + n \\ \delta_n &= \delta_0 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2 \end{aligned}$$

# Normal model: Example



# Normal model: Example




$$p(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)$$

# Posterior expectations

- Up until now, we have been able to get a closed form solution for the posterior
- This allows us to obtain many quantities related to the posterior such as mean, variance or probabilities (for credibility intervals)
- Most of these quantities of interest can be expressed using expectations over the posterior $p(\theta|y)$:
  - Mean: $\mathrm{E}[\theta|y] = \int \theta p(\theta|y) d\theta$
  - Variance: $\mathrm{E}[(\theta - \mathrm{E}[\theta|y])^2|y] = \int (\theta - \mathrm{E}[\theta|y])^2 p(\theta|y) d\theta$
  - Probability of $\mathcal{A}$: $\Pr(\theta \in \mathcal{A}|y) = \int 1(\theta \in \mathcal{A}) p(\theta|y) d\theta$
  - Marginal of $\theta_1$: $p(\theta_1|y) = \int p(\theta_1|\theta_2, y) p(\theta_2|y) d\theta_2$

# Posterior expectations

- In general, this means our objects of interest will be quantities of the form

$$\mathrm{E}_{p(\theta|y)}[f(\theta)] = \int f(\theta)p(\theta|y)d\theta$$

for some function $f(\theta)$ where

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

- We can easily evaluate $p(y|\theta)p(\theta)$ for any $\theta$, but the integral $\int p(y|\theta)p(\theta)d\theta$ is usually difficult
- We will now focus on ways to evaluate these expectations when we lack closed form solutions for the posterior and without requiring that we evaluate $\int p(y|\theta)p(\theta)d\theta$

# Numerical Integration

- Need to numerically evaluate expectations (integrals) of the form

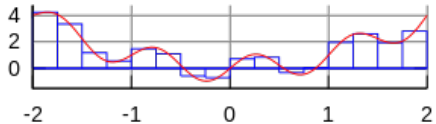
$$\mathrm{E}_{p(\theta|y)}[f(\theta)] = \int f(\theta)p(\theta|y)d\theta$$

- We integrate over possible values of $\theta$ so dimensionality of $\theta$ is crucial
- Many methods exist for numerically approximating posterior expectations:
  1. Conjugate priors and analytic solutions (BDA Ch. 1-5)
  2. Grid integration and other quadrature rules (BDA Ch. 3, 10)
  3. Monte Carlo integration, rejection and importance sampling (BDA Ch. 10)
  4. Markov Chain Monte Carlo (BDA Ch. 11-12)
  5. Approximate Bayesian Computation (ABC): Distributional approximations (Laplace, VB) (BDA Ch. 4, 13) and minimum distance

# Quadrature integration

- The simplest quadrature integration is grid integration: Evaluate function in a grid and compute

$$\mathrm{E}[f(\theta)] \approx \sum_{g=1}^{G} w_{\mathrm{cell}}^{(g)} f(\theta(g)),$$

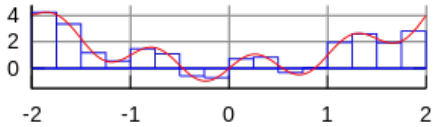



where $w_{\mathrm{cell}}^{(g)}$ is the normalized probability of a grid cell $g$, and $\theta^{(g)}$ are center locations of grid cells

# Quadrature integration

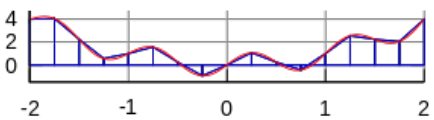
- The simplest quadrature integration is grid integration: Evaluate function in a grid and compute

$$\mathrm{E}[f(\theta)] \approx \sum_{g=1}^{G} w_{\mathrm{cell}}^{(g)} f(\theta(g)),$$




where $w_{\mathrm{cell}}^{(g)}$ is the normalized probability of a grid cell $g$, and $\theta^{(g)}$ are center locations of grid cells

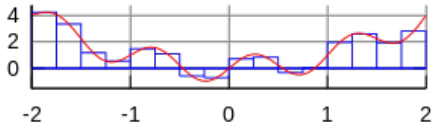
- In 1D further variations with smaller error, e.g. trapezoid



# Quadrature integration

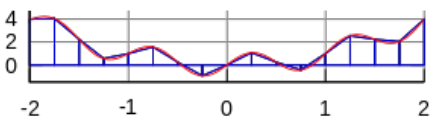
- The simplest quadrature integration is grid integration: Evaluate function in a grid and compute

$$\mathrm{E}[f(\theta)] \approx \sum_{g=1}^{G} w_{\mathrm{cell}}^{(g)} f(\theta(g)),$$




where $w_{\mathrm{cell}}^{(g)}$ is the normalized probability of a grid cell $g$, and $\theta^{(g)}$ are center locations of grid cells

- In 1D further variations with smaller error, e.g. trapezoid




- In 2D and higher: nested quadrature and product rules

# Monte Carlo

- If we are able to draw samples $(\theta^{(1)}, \ldots, \theta^{(S)})$ from $p(\theta|y)$, then by the law of large numbers

$$\frac{1}{S}\sum_{s=1}^{S} f(\theta^{(s)}) \xrightarrow{p} \mathrm{E}_{p(\theta|y)}[f(\theta)]$$

- Use these draws, for example,
  - to compute means, deviations, quantiles
  - to draw histograms
  - to marginalize

# Monte Carlo vs. deterministic

- Monte Carlo = simulation methods
  - evaluation points are selected stochastically (randomly)
  - performance depends on amount of simulations and many may be needed for accurate answers
- Deterministic methods (e.g. grid)
  - evaluation points are selected by some deterministic rule
  - good deterministic methods converge faster than simulation ones (they need less function evaluations)

# Direct Sampling

- Idea: If $p(\theta|y)$ is of a known form, simply draw samples of $\theta$ from it
- Produces independent draws
- Problem: restricted to limited set of models
- Most models where direct sampling is available are due to conjugate priors and analytical solutions

# Grid sampling and curse of dimensionality

- Number of grid points increases exponentially with the dimensions of $\theta$
- if we don't know beforehand where the posterior mass is
  - need to choose wide box for the grid
  - need to have enough grid points to get some of them where essential mass is
- e.g. 10 parameters with 50 or 1000 grid points per dimension
  - $50^{10} \approx$ 1e17 grid points
  - $1000^{10} \approx$ 1e30 grid points
- If you can compute likelihood about 20 million times per second
  - Evaluation in 1e17 grid points would take 150 years
  - Evaluation in 1e30 grid points would take 1500 billion years

# Indirect sampling

- Rejection sampling (Lab)
  - draw directly from a proposal distribution, reject some draws, remaining draws are independent draws from the target distribution

# Indirect sampling

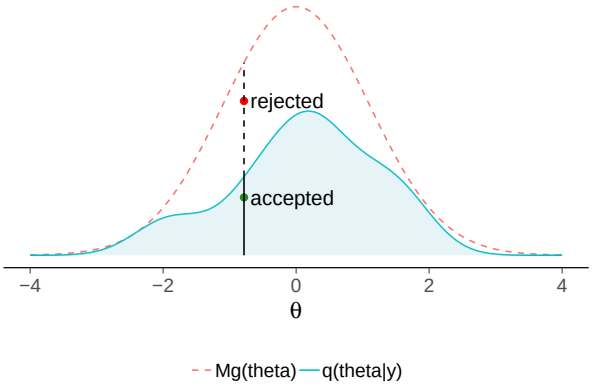
- Rejection sampling (Lab)
  - draw directly from a proposal distribution, reject some draws, remaining draws are independent draws from the target distribution
- Importance sampling (Lab)
  - draw directly from a proposal distribution, weight the draws

# Indirect sampling

- Rejection sampling (Lab)
  - draw directly from a proposal distribution, reject some draws, remaining draws are independent draws from the target distribution
- Importance sampling (Lab)
  - draw directly from a proposal distribution, weight the draws
- Markov chain Monte Carlo (next week)
  - draw directly from a transition distribution forming a Markov chain, draws are dependent draws from the target distribution

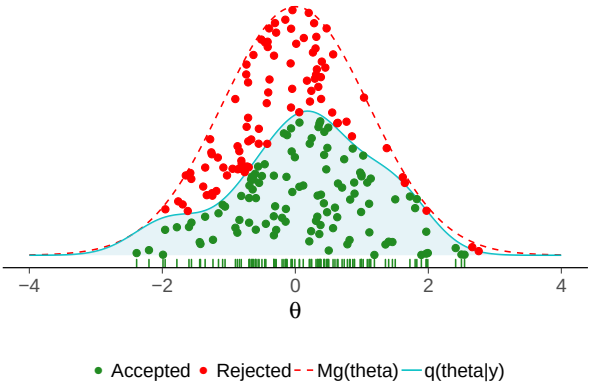
# Rejection sampling

- Proposal forms envelope over the target distribution $q(\theta|y)/Mg(\theta) \leq 1$
- Draw from the proposal and accept with probability $q(\theta|y)/Mg(\theta)$



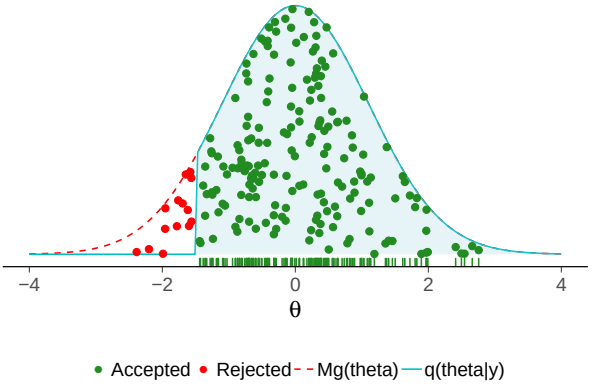
# Rejection sampling

- Proposal forms envelope over the target distribution $q(\theta|y)/Mg(\theta) \leq 1$
- Draw from the proposal and accept with probability $q(\theta|y)/Mg(\theta)$



# Rejection sampling

- Proposal forms envelope over the target distribution $q(\theta|y)/Mg(\theta) \leq 1$
- Draw from the proposal and accept with probability $q(\theta|y)/Mg(\theta)$
- Common for truncated distributions



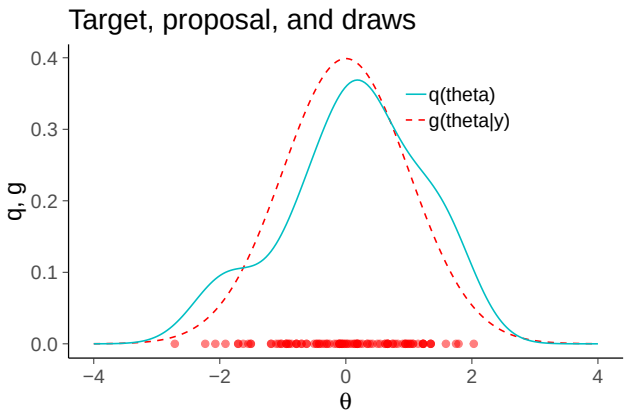
# Rejection sampling

The number of accepted draws is the effective sample size

- with bad proposal distribution may require a lot of trials
- selection of good proposal gets very difficult when the number of dimensions increase
- for truncated distributions, effective sample size depends on the area of truncated region

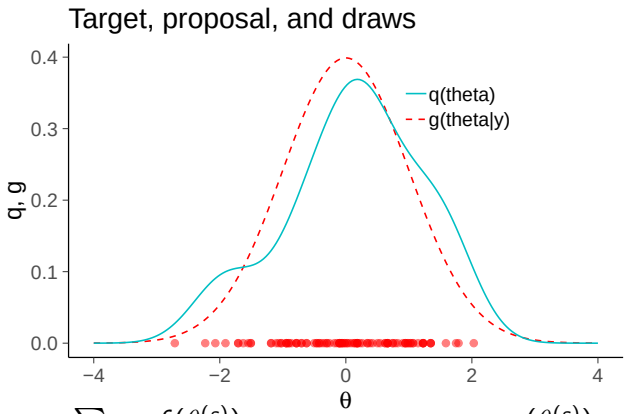
# Importance sampling

Proposal does not need to have a higher value everywhere



# Importance sampling

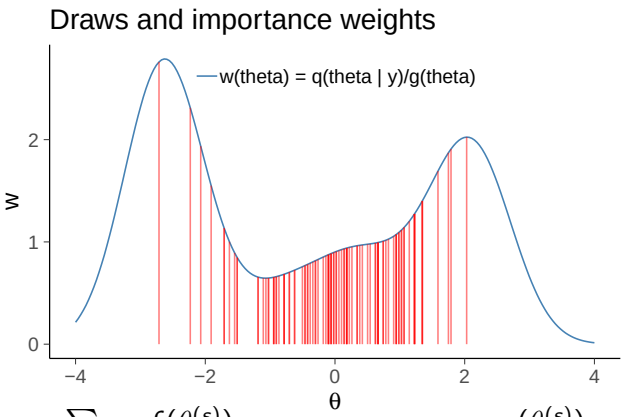
Proposal does not need to have a higher value everywhere




$$\mathrm{E}[f(\theta)] \approx \frac{\sum_s w_s f(\theta^{(s)})}{\sum_s w_s}, \qquad \text{where} \quad w_s = \frac{q(\theta^{(s)})}{g(\theta^{(s)})}$$

# Importance sampling

Proposal does not need to have a higher value everywhere




$$\mathrm{E}[f(\theta)] \approx \frac{\sum_s w_s f(\theta^{(s)})}{\sum_s w_s}, \qquad \text{where} \quad w_s = \frac{q(\theta^{(s)})}{g(\theta^{(s)})}$$

# Importance sampling

- Resampling using normalized importance weights can be used to pick a smaller number of draws with uniform weights
- Selection of good proposal gets more difficult when the number of dimensions increase
- Often used to correct distributional approximations
- Variation of the weights affect the effective sample size
  - if single weight dominates, we have effectively one sample
  - if weights are equal, we have effectively $S$ draws
- Central limit theorem holds only if variance of the weight distribution is finite

# Markov chain Monte Carlo (MCMC)

- Pros
  - Markov chain goes where most of the posterior mass is
  - Certain MCMC methods scale well to high dimensions
- Cons
  - Draws are dependent (affects how many draws are needed)
  - Convergence in practical time is not guaranteed
- MCMC methods in this course
  - Gibbs: iterative conditional sampling
  - Metropolis: random walk in joint distribution
  - Dynamic Hamiltonian Monte Carlo: "state-of-the-art" used in Stan