# Applied Bayesian Modeling module 5: **Models with more than 1 parameters, sampling**

Leontine Alkema, lalkema@umass.edu

Fall 2022

# Models with more than 1 parameter

- We discussed Bayesian inference for $\mu$ when $y_i|\mu, \sigma \sim N(\mu, \sigma^2)$, with $\sigma$ known.
- What if $\sigma$ is unknown?
- To discuss:
  - Bayes' rule for inference for $> 1$ parameter.
  - Using samples to do inference

## Bayes' rule for more than 1 parameter

- Goal: estimate $(\mu, \sigma)$ when $y_i|\mu, \sigma \sim N(\mu, \sigma^2)$.
- If we put a joint prior $p(\mu, \sigma)$ on the parameters, Bayes' rule tells us how to get the joint posterior distribution:

$$p(\mu, \sigma|\boldsymbol{y}) = \frac{p(\mu, \sigma)p(\boldsymbol{y}|\mu, \sigma^2)}{p(\boldsymbol{y})} \propto p(\mu, \sigma)p(\boldsymbol{y}|\mu, \sigma^2).$$

- And if inference about $\mu$ is our goal, we can get the *marginal* posterior distribution $p(\mu|\boldsymbol{y}) = \int_\sigma p(\mu, \sigma|\boldsymbol{y})d\sigma$.
- That's good news!

...

# The bad news

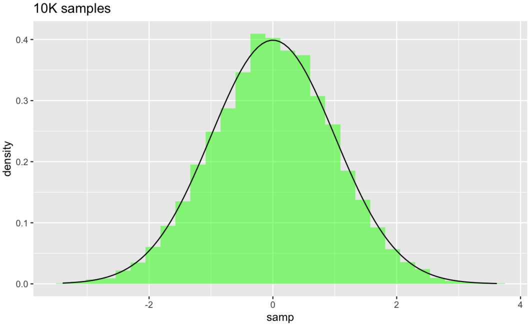
- The good news so far: Bayes' rule tells us how to get a joint posterior distribution for more than 1 parameter

$$p(\mu, \sigma|\boldsymbol{y}) = \frac{p(\mu, \sigma)p(\boldsymbol{y}|\mu, \sigma^2)}{p(\boldsymbol{y})} \propto p(\mu, \sigma)p(\boldsymbol{y}|\mu, \sigma^2).$$

- The bad news:
  Common choices of priors do not result in a closed-form expression for $p(\boldsymbol{y})$, hence we don't get a closed-form expression for $p(\mu, \sigma|\boldsymbol{y})$ and/or $p(\mu|\boldsymbol{y})$.
  - In closed form $=$ as a mathematical expression that can be evaluated exactly in a finite number of operations
- Back to more good news:
  No problem if we do not have a closed-form expressions for a pdf, as long as we can get a sample from the distribution of interest, we can do inference!

# Simulation-based inference

- The general idea in simulation-based inference:
  We can make inference about a random variable $\mu$, using a sample $\{\mu^{(1)}, \ldots, \mu^{(S)}\}$ from its probability distribution.
  This is called a Monte Carlo (MC) approximation.
- Example: learn about $\mu$ using samples, illustrated for $\mu \sim N(0, 1)$



# Monte Carlo approximation for the mean of a random variable

- If we know that $\mu \sim N(0,1)$, then we can calculate $E(Y) = \int \mu p(\mu) d\mu = 0$.
- If we do NOT know the probability distribution but we HAVE a sample from it, an MC approximation is given by the sample mean

```
set.seed(1234)
samp <- rnorm(10000, 0, 1)
mean(samp)
[1] 0.006115893
```

- Details: Why can we use a sample mean as an approximation to the mean of a random variable?
  - Based on the law of large numbers we know that: $\frac{1}{S}\sum_{s=1}^{S} \mu^{(s)} \to E(\mu)$ as sample size $S \to \infty$.
  - The error in the MC approximation for $E(\mu)$ goes to zero as the sample size increases because $Var(\frac{1}{S}\sum_{s=1}^{S} \mu^{(s)}) = \frac{Var(\mu)}{S} \to 0$.

# Monte Carlo approximation for other outcomes of interest

- Just about any aspect of the distribution of $\mu$ can be approximated arbitrarily exactly with a large enough Monte Carlo sample, e.g.
  - the $\alpha$-percentile of $\{\mu^{(1)}, \ldots, \mu^{(S)}\} \to$ the $\alpha$-percentile of the distribution, e.g.
    the 2.5th percentile of $\{\mu^{(1)}, \ldots, \mu^{(S)}\} \to$ 2.5th percentile of $p(\mu)$
  - Details: We can approximate $Pr(\mu \leq x)$ for any constant $x$ by the proportion of samples for which $\mu \leq x$, because

$$1/S \sum_{s=1}^{S} I(\mu^{(s)} \leq x) \to Pr(\mu \leq x).$$

- So we can approximate the mean, median, and credible intervals for $\mu$ using a sample

# Monte Carlo approximation: further use

- With a simulation, it also becomes very easy to analyze the distributions of any function of 1 or more random variables, e.g.
  - the distribution of $1/\mu$ by using samples $1/\mu^{(s)}$,
  - the distribution of the ratio $\mu_1/\mu_2$ can be studied using the ratio of the samples $\mu_1^{(s)}/\mu_2^{(s)}$ with $(\mu_1^{(s)}, \mu_2^{(s)}) \sim p(\mu_1, \mu_2)$,
- Samples from mariginal distributions may be obtained from samples from joint distributions, e.g.
  - if $(\mu_1^{(s)}, \mu_2^{(s)}) \sim p(\mu_1, \mu_2)$, then $\mu_1^{(s)} \sim p(\mu_1)$

Conclusion: good news! We don't need $p(\mu|\boldsymbol{y})$ in closed form as long as we can obtain samples from it!

# Back to the example

- Goal: estimate $(\mu, \sigma)$ when $y_i|\mu, \sigma \sim N(\mu, \sigma^2)$.
- Problem: For common choices of the priors on $\mu$ and $\sigma$, there is no closed-form expression for $p(\mu|\boldsymbol{y})$.
- Solution: let's obtain posterior samples $\mu^{(1)}, \ldots, \mu^{(S)} \sim p(\mu|\boldsymbol{y})$
- How?
  - "MCMC-goodness comes to the rescue!
  - Samples can be obtained through an MCMC algorithm
- Next couple of slides: brief intro into MCMC, details to follow in next modules

# Markov Chain Monte Carlo (MCMC) algorithm

- Let $\boldsymbol{\phi}$ = parameter vector of interest, i.e. $\boldsymbol{\phi} = (\mu, \sigma)$.
- Goal: obtain samples $\boldsymbol{\phi}^{(s)}$ from the target distribution, here $p(\boldsymbol{\phi}|\boldsymbol{y})$
- MCMC approach:
  - get some initial value $\boldsymbol{\phi}^{(1)}$ and create a sequence $\boldsymbol{\phi}^{(1)}, \boldsymbol{\phi}^{(2)}, \ldots$
  - such that for some large $s$,
  - $\boldsymbol{\phi}^{(s)}$ is a draw from the target distribution
- In MCMC, $\boldsymbol{\phi}^{(s)}$ depends on $\boldsymbol{\phi}^{(s-1)}, \boldsymbol{\phi}^{(s-2)}, \ldots, \boldsymbol{\phi}^{(1)}$ only through $\boldsymbol{\phi}^{(s-1)}$. This is called the Markov property, and so the sequence is called a **Markov chain**.
- We approximate quantities of interest, e.g. $E(\mu|\boldsymbol{y})$, using resulting samples, which adds in the **Monte Carlo** part.
- You may/should wonder...

## How to sample the $\phi^{(s)}$ in an MCMC algorithm?

- General idea
  - Propose a new value $\phi^{(s)}$
  - Accept or reject (set $\phi^{(s)} = \phi^{(s-1)}$) based on target distribution
- How? Different strategies, aimed at finding efficient proposals, i.e.
  - Metropolis hasting,
  - Class of proposal distributions that leads to Hamilton Monte Carlo.

  Generally, there are MCMC parameters that need to be tuned to help the chain converge to the target distribution/sample most efficiently.
- We will use software Stan for this, with built-in samplers and automated tuning of MCMC parameters (mostly).
- More in next modules on that, and what to check before working with outputs.

## Summary, for parameter vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)$:

- Bayes rule when estimating $\boldsymbol{\mu}$:

  $$p(\boldsymbol{\mu}|\boldsymbol{y}) = p(\boldsymbol{\mu}, \boldsymbol{y})/p(\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\mu})p(\boldsymbol{\mu})/p(\boldsymbol{y}),$$

  with the marginal posterior for just one parameter is given by:
  $p(\mu_1|\boldsymbol{y}) = \int_{\mu'_2} \cdots \int_{\mu'_p} p(\mu_1, \mu'_2, \ldots, \mu'_p|\boldsymbol{y}) d\mu'_2 \cdots d\mu'_p$.
- Often, we don't have a closed-form expression for $p(\boldsymbol{\mu}|\boldsymbol{y})$. Then sampling comes to the rescue:
  - We can make inference about $\boldsymbol{\mu}$ using a sample $\{\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(S)}\} \sim p(\boldsymbol{\mu}|\boldsymbol{y})$. This is called a Monte Carlo (MC) approximation.
  - We can report any summary we'd like, e.g. posterior mean (sample mean), posterior median or other percentiles (sample percentiles).
  - We can sample from posterior distributions using an MCMC algorithm; details to come up next!
- Next lab: get going with Stan, fit a simple model using the package "brms" (which uses Stan).