STA2201H Methods of Applied Statistics II

Monica Alexander

Week 3: Logistic regression, Intro to Bayesian Inference

Announcements

- We will be online until after reading week.
- This Friday at 12pm, Toronto Data Workshop: Ashok Chaurasia, University of Waterloo, 'Multiple Imputation: Old and New Combining Rules for Statistical Inference'.
- This Friday at 1pm, Formal Demography Working Group

Data in context, applied statistics, and answering hard questions

Consider the following model

$$y_i \sim (\mu_i, \sigma^2)$$

Say y_i is birthweight, and we are interested in understanding what factors influence this outcome

- How can we model μ_i?
- What if the data come from a survey?
- What if the data are at the national level and we have multiple surveys?

Data in context, applied statistics, and answering hard questions

$$y_i \sim (\mu_i, \sigma^2)$$

- You can't implement a good model without understanding your data
- You can't understand your data (and the data generating process) without understanding context

Data in context, applied statistics, and answering hard questions

- Applied statistics necessarily requires engagement with the data
- Without context, data lose their meaning
- Asking difficult questions is part of doing science

The University's policy on Academic Freedom is here (Article 5)

Logistic regression

Example: Migration to Florida

- Data from 2019 ACS
- Outcome of interest: ‘moved to Florida in last year (yes/no)’ for people residing in Florida
- Other variables: age, employment status, education

What the data look like:

serial	moved_into_FL	age	graduated_high_school	empstat
271269	0	33	0	not in labor force
271270	0	56	1	not in labor force
271271	0	33	1	not in labor force
271272	0	19	0	employed
271273	0	48	0	not in labor force
271274	0	59	0	not in labor force

Migration outcomes

Consider the model

$$Y_i \sim \text{Bernoulli}(\pi_i)$$

$$\text{logit}\pi_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{emp}_i + \beta_3 \text{school}_i$$

where school_i whether or not respondent graduated high school.

What could we use this model for? (i.e. what questions could we ask?)

Estimation in R

```r
mod <- glm(moved_into_FL ~ age + empstat + graduated_high_school, data = d, family = "binomial")
summary(mod)
```

```
##
## Call:
## glm(formula = moved_into_FL ~ age + empstat + graduated_high_school,
##     family = "binomial", data = d)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.4793  -0.2970  -0.2620  -0.2388   2.8933
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -2.6325155  0.0376894 -69.848  < 2e-16 ***
## age                        -0.0170847  0.0006533 -26.152  < 2e-16 ***
## empstatunemployed           0.6854172  0.0620935  11.038  < 2e-16 ***
## empstatnot in labor force   0.3589597  0.0267108  13.439  < 2e-16 ***
## graduated_high_school       0.1143906  0.0264405   4.326 1.52e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 57241  on 175754  degrees of freedom
## Residual deviance: 56414  on 175750  degrees of freedom
## AIC: 56424
##
## Number of Fisher Scoring iterations: 6
```

Interpretation

```
coef(mod)
```

```
##               (Intercept)                       age          empstatunemployed
##               -2.63251546               -0.01708474                 0.68541725
## empstatnot in labor force      graduated_high_school
##                0.35895974                0.11439060
```

```
exp(coef(mod))
```

```
##               (Intercept)                       age          empstatunemployed
##                0.07189738                0.98306038                 1.98459973
## empstatnot in labor force      graduated_high_school
##                1.43183915                1.12118998
```

Questions

- What is the probability that a Florida resident moved there last year, if they are aged 25, employed, and didn't graduate high school?

```
estimated_log_odds <- coef(mod)[1] + coef(mod)[2]*25
exp(estimated_log_odds)/(1+exp(estimated_log_odds))
```

```
## (Intercept)
##  0.04480337
```

Questions

- Assume we don't observe people living in Tampa Bay. Could we use this model to predict the likelihood of having migrated for Tampa Bay residents?
- Can we use this model to estimate the impact of education on migration?

What are some potential issues with this analysis?

$$\text{logit}\pi_i = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{emp}_i + \beta_3 \text{school}_i$$

Model issues

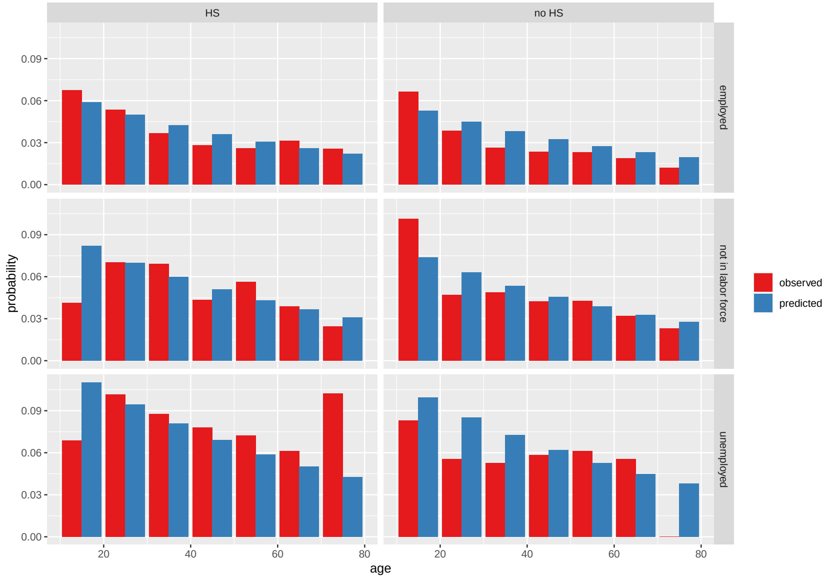
- Omitted variable bias
- Model mis-specification
- Model underfit or overfit
- Multicollinearity

Tools (for now)

- EDA!
- Likelihood ratio tests
- Wald tests
- Assessing predictions/residuals graphically (harder with binary variables)

A good way of assessing model fit

Look at predicted v actual proportions by groups



Issues with causal questions

Consider the situations where we are interested in the impact of education on migration outcomes.

```
summary(mod)
```

```
##
## Call:
## glm(formula = moved_into_FL ~ age + empstat + graduated_high_school,
##     family = "binomial", data = d)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.4793  -0.2970  -0.2620  -0.2388   2.8933
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -2.6325155  0.0376894 -69.848  < 2e-16 ***
## age                       -0.0170847  0.0006533 -26.152  < 2e-16 ***
## empstatunemployed          0.6854172  0.0620935  11.038  < 2e-16 ***
## empstatnot in labor force  0.3589597  0.0267108  13.439  < 2e-16 ***
## graduated_high_school      0.1143906  0.0264405   4.326 1.52e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 57241  on 175754  degrees of freedom
## Residual deviance: 56414  on 175750  degrees of freedom
## AIC: 56424
##
## Number of Fisher Scoring iterations: 6
```

Issues with causal questions

Consider the situations where we are interested in the impact of education on migration outcomes. This is a causal question. What are some issues that may arise?

Issues with causal questions

- Confounders
  - urbanity
- Colliders (e.g. non-reponse bias)
  - Education and migration both influence survey response
  - Conditioning on survey response creates a noncausal association between education and migration

Data issues

- Non-representative samples
- Non-response (complete survey or specific questions)
- Measurement error

Introduction to Bayesian Inference

Readings

- Gelman, Carlin, Stern, Dunson, Ventari and Rubin (2013). Bayesian Data Analysis (Third Edition) Chapman and Hall/CRC
  - Aki's slides on BDA are useful: https://github.com/avehtari/BDA_course_Aalto
- Gelman and Hill (2006). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University
- Hoff (2010). A first course in Bayesian statistical methods
- If interested in something a bit more philosophical: Stark (2015). Constraints versus priors. SIAM/ASA Journal on Uncertainty Quantification, 3(1), 586-598.

Back to linear regression

We model the relationship between the (potentially transformed) data and covariates as a linear regression model

$$g(y_i) = {x_i}^T \beta + \epsilon_i$$

- Previously, you have probably written down the likelihood and found the MLE estimate(s) for β. Look something like

  $$\hat{\beta} = (X^T W X)^{-1} X^T W z$$

  where $z = f(y, X, \beta, g)$ and for usual linear regression, the weights W are the identity and $z = y$.
- Once we have $\hat{\beta}$s, can assume asymptotic normality and do some inference

What are we doing here

This type of classical inference (= **frequentist** inference) has an underlying probabilistic framework:

- The data y are random
- The estimator $\hat{\beta}$ is a function of the data
- We can then make probability statements about how often the true value is within some interval around the estimator.
- So we are always making probabilistic statements about the true value of β and how uncertain we are as a function of the data

Let's ask a different question

Which values of β are consistent with the data we have observed?

"*Given* the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of it happening in a single trial lies somewhere between any two degrees of probability that can be named."

Bayesian versus frequentist

Frequentist

- Parameter(s) θ is a fixed but unknown quantity
- Probability: to describe the relative frequency of an outcome in an infinitely repeatable but unpredictable experiment
- Uncertainties typically involve expectations with respect to the distribution of the data, holding the parameter fixed

Bayesian

- Parameter(s) θ is a random variable
- Probability statements reflect a state of knowledge
- Uncertainties typically involve expectations with respect to the distribution of the parameter, holding the data fixed

Bayesian inference

Bayesian inference

The process of learning via Bayes rule.

Example: breast cancer screening (using mammograms) in Germany. Imagine we know

- The probability an asymptomatic woman has breast cancer is 0.8%.
- If she has breast cancer, the probability is 90% that she has a positive mammogram
- If she does not have breast cancer, the probability is 7% that she still has a positive mammogram.

Suppose a woman has a positive mammogram: What is the probability she actually has breast cancer?

Breast cancer

Use Bayes rule for events:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Let

- C be the cancer outcome (=1 if cancer, 0 otherwise)
- M be the mammogram outcome (=1 if mammogram is positive, 0 otherwise)

"Suppose a woman has a positive mammogram: What is the probability she actually has breast cancer?"

A somewhat famous example because physicians had no idea what the answer should be.

We want to know $P(C = 1|M = 1)$.

Breast cancer

We want to know $P(C = 1|M = 1)$.

- $P(C = 1) = 0.008$.
- $P(M = 1|C = 1) = 0.9$.
- $P(M = 1|C = 0) = 0.07$.
- so $P(M = 1) =?$

Use Bayes rule, get $P(C = 1|M = 1) = 9.4\%$.

What did we do? Updated **prior** probability $P(C = 1)$ based on observing **data** (mammograms) to get the **posterior** probability $P(C = 1|M = 1)$.

Bayesian inference about parameters

Happiness example

Hoff (Chapter 3):

- Each female aged 65+ in 1998 General Social Survey was asked about being happy.
- Data: Out of $n = 129$ women, $y = 118$ women (91%) reported being happy.
- What is $\theta =$ the proportion of 65+ women who are happy?
- Goal: inference about $\theta =$ happiness parameter.

Happiness example

What's our usual approach? (frequentist)

1. Relate data to parameter of interest through a likelihood function, e.g. assume $Y|\theta \sim Bin(n, \theta)$ where y is the number of women who report to be happy out of the sample of n women.
2. Maximum likelihood estimate: Find a point estimate θ that maximizes the likelihood function ($\hat{\theta} = 0.91$)
3. Construct a confidence interval for θ (CI: [0.87, 0.96])
4. Interpretation of frequentist CI: If repeated samples were taken and the 95% confidence interval was computed for each sample, 95% of the intervals would contain the population mean.

Happiness example

The Bayesian approach:

- Also assume a likelihood, as before $Y|\theta \sim Bin(n, \theta)$

But now we proceed differently. In Bayesian inference, unknown parameters (like θ) are considered **random variables**. This means information/knowledge about these random variables can be summarized using probability distributions.

- Have existing knowledge/info about θ, summarized by the prior probability distribution
- Observe some data that gives more info about θ
- Update our previous knowledge to obtain the posterior distribution using Bayes' rule

Happiness example

The Bayesian approach:

1. Also assume a likelihood $p(y|\theta)$, as before $Y|\theta \sim Bin(n, \theta)$
2. Set a prior distribution for θ, $p(\theta)$
3. Use Bayes rule to update the prior into the posterior distribution
$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$
4. Use the posterior to provide summaries of interest, e.g. point estimates and uncertainty intervals, called credible intervals.

Happiness example

1. Likelihood is $Y|\theta \sim Bin(n, \theta)$ so

$$p(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$$

2. Now we need to pick a prior $p(\theta)$

- Suppose any outcome between 0 and 1 for θ is equally likely, what prior can be used to describe these beliefs?
- $\theta \sim U(0, 1)$ so $p(\theta) = 1$

3. Now we calculate the posterior distribution

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

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

In the happiness case,

$$p(\theta|y) = \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\int_0^1 \binom{n}{y}\theta'^y(1-\theta')^{n-y}d\theta'} = \frac{1}{Z}\theta^y(1-\theta)^{n-y}$$

where

$$Z = \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)}$$

So posterior is

$$\theta|y \sim \text{Beta}(y+1, n-y+1)$$

Up to a constant

To recognize the posterior as a Beta distribution, it would have been sufficient to consider only the terms that include θ

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

i.e.

$$p(\theta|y) \propto \theta^{y}(1-\theta)^{n-y}$$

because $p(\theta|y)$ is a pdf so must integrate to one. So the marginal distribution $p(y)$ is just a scaling factor.

Inference about θ based on posterior distribution

Bayesian point estimates are often given by:

- The posterior mean $E(\theta|y)$
- The posterior median θ^* $P(\theta < \theta^*|y) = 0.5$.

Uncertainty is quantified with credible intervals (CIs), e.g. for 95% CIs:

- An interval is called a 95% Bayesian CI if the posterior probability that θ is contained in the interval is 0.95.
- More formally a $1-\alpha$ credible interval for θ is an interval C_n satisfying $P(\theta \in C_n|Y_1, \ldots, Y_n) = 1-\alpha$.
  - a probability statement about θ, not C_n.

Bayesian credible intervals

An interval is called a 95% Bayesian CI if the posterior probability that θ is contained in the interval is 0.95.

- More formally a $1-\alpha$ credible interval for θ is an interval C_n satisfying $P(\theta \in C_n | Y_1, \ldots, Y_n) = 1-\alpha$.
- a probability statement about θ (given the data), not C_n.
- "the probability that θ is in C_n given the data is 95%"

This interpretation differs from a frequentist CI; it is a statement about the information about the location of θ.

- c.f. confidence interval: a $1-\alpha$ confidence interval for θ is an interval C_n satisfying $P(\theta \ni C_n) \geq 1-\alpha$
- a probability statement about C_n, not θ
- "if I repeat the experiment over and over, the interval will contain the parameter 95% of the time."

Bayesian credible intervals

C_n is not uniquely defined. Interval options:

- Quantile-based Bayesian $100(1-\alpha)\%$ CIs are used, which are given by posterior quantiles $(\theta_{\alpha/2}, \theta_{1-\alpha/2})$, with $P(\theta < \theta_{\alpha/2}|y) = P(\theta > \theta_{1-\alpha/2}|y) = \alpha/2$. (focus here)
- Highest posterior density (HPD) intervals (see here for more details).

Happiness findings

Bayesian estimates: Mean

```
## [1] 0.91
```

95% Credible interval:

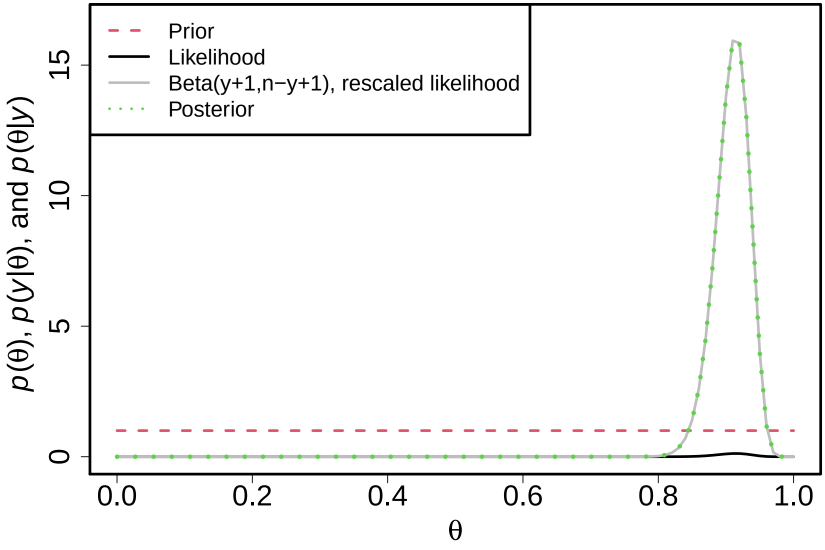
```
## [1] 0.85 0.95
```

Frequentist estimates (mean, 95% CI)

```
## [1] 0.91
```

```
## [1] 0.87
```

```
## [1] 0.96
```



Prior
Likelihood
Beta(y+1,n−y+1), rescaled likelihood
Posterior
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
θ
p(θ), p(y|θ), and p(θ|y)

Conjugate priors

Note that $\theta \sim U(0, 1)$ is the same as $\theta \sim \mathit{Beta}(1, 1)$.

For the binomial likelihood, a Beta prior results in a Beta posterior distribution: we say that the beta prior is **conjugate** for the binomial likelihood.

More generally, for a certain likelihood, a prior distribution which results in a posterior distribution of the same form is called a **conjugate** prior distribution.

Priors

Different types of priors

BDA Chapter 2

- Conjugate prior
- Noninformative prior
- Proper and improper prior
- Weakly informative prior
- Informative prior

Conjugate priors

- Prior and posterior have the same form
- only for exponential family distributions (plus for some irregular cases)
- Used to be important for computational reasons

e.g beta for binomial. What's the interpretation of a Beta(1,1) prior?

Noninformative prior, proper and improper prior

- Vague, flat, diffuse of noninformative
- "let the data speak for themselves"

But flat is not non-informative!

Proper prior: $\int p(\theta) = 1$

Improper prior: doesn't have finite integral (but the posterior can still sometimes be proper)

- e.g. The uniform distribution on an infinite interval (i.e., a half-line or the entire real line).

Weakly informative priors

- Quite often there's at least some knowledge about the scale
- The idea is that the prior rules out unreasonable parameter values but is not so strong as to rule out values that might make sense
- Weakly informative priors produce computationally better behaving posteriors
- Generic weakly informative prior: $N(0, 1)$
- Good example in the Gabry et al paper on air pollution
- More on this in a couple of lectures

Informative priors

Prior distributions giving numerical information that is crucial to estimation of the model. Information might come from a literature review or explicitly from an earlier data analysis.

- Example from Gelman (linked): Mass of liver as a fraction of lean body mass is known to vary very little.
- E.g. Gompertz models for mortality: can only have a restricted range on α and β that lead to plausible values of life expectancy

Bias-variance tradeoff

- Effect of incorrect priors: Introduce bias, but often still produce smaller estimation error because the variance is reduced
- Misleading certainty in results?

How to choose?

Some good practical advice: https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations

- if you do have prior info, include it!
- make sure it has appropriate range (e.g. prior on variance needs to have positive support)
- prior predictive checks with simulated data
- check sensitivity of model findings to model choice

More on this later.

More than one parameter

More than one parameter

What if the data model (likelihood function) includes more than 1 unknown parameter, e.g. do inference for μ if

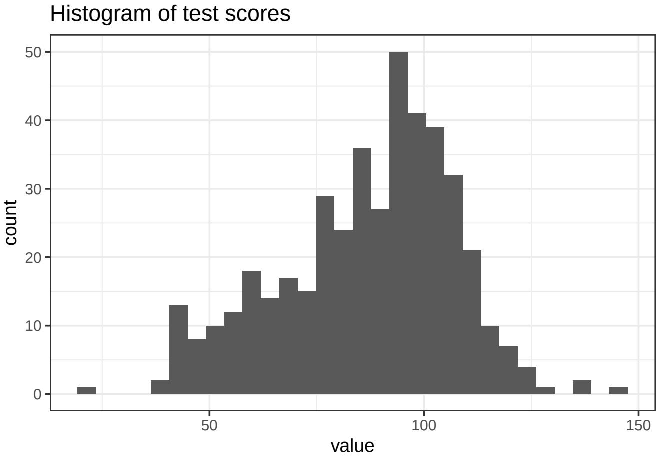
$$y_i \sim N(\mu, \sigma^2)$$

What do we want? $p(\mu, \sigma | \mathbf{y})$. If only mean is of interest, then want $p(\mu | \mathbf{y})$.

Example: kid's test scores

Gelman-Hill Chapter 3 Outcome of interest: cognitive tests scores for 3-4 year old kids. Denote the unknown mean test score by μ, and observed test score by y_i for kid i, with $i = 1, \ldots, n$.

Goal: estimate μ.



Example: kid's test scores

Let's assume Normal likelihood

$$y_i \sim N(\mu, \sigma^2)$$

- If we put a joint prior $p(\mu, \sigma)$ on the parameters, Bayes' rule tells us how to get the joint posterior distribution:

$$p(\mu, \sigma|\mathbf{y}) = \frac{p(\mathbf{y}|\mu, \sigma)p(\mu, \sigma)}{p(\mathbf{y})}$$

- And if inference about μ is our goal, we can get the marginal posterior distribution

$$p(\mu|\mathbf{y}) = \int_{\sigma} p(\mu, \sigma|\mathbf{y})d\sigma$$

Example: kid's test scores

- What priors to set for μ and σ^2? Let's assume that μ and σ^2 are independent a priori, $p(\mu, \sigma) = p(\mu)p(\sigma)$, and use

$$\mu \sim N(\mu_0, \sigma^2_{\mu_0})$$

and

$$1/\sigma^2 \sim Gamma(\nu_0/2, \nu_0/2 \cdot \sigma_0^2)$$

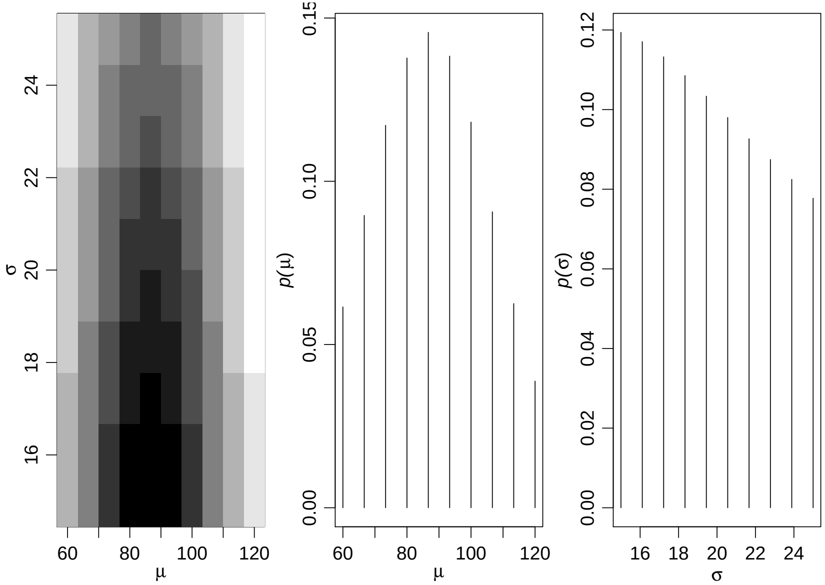
For illustrative purposes, will set **hyperparameters** to be $\mu_0 = 86.8$, $\sigma_{\mu_0} = \sigma_0 = 20.4$ and $\nu_0 = 1$.

Let's start with a discrete approximation

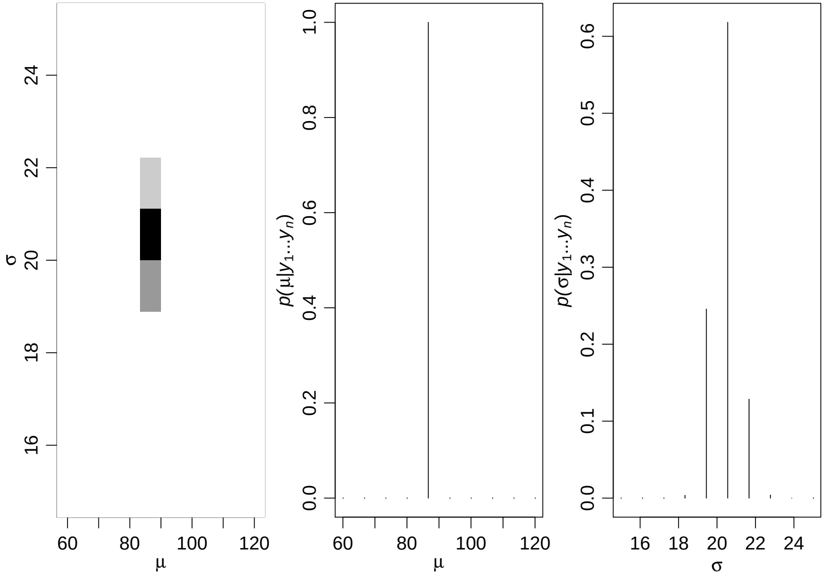
For illustrative purposes, start with a discrete approximation to these priors.

- E.g. use discrete grid of values for μ, and set $p(\mu) = f(\mu)/\sum f(\mu)$ where $f(.)$ is given by the Normal pdf for μ.
- Can use these to calculate discrete likelihood and thus a discrete approximation to the posterior

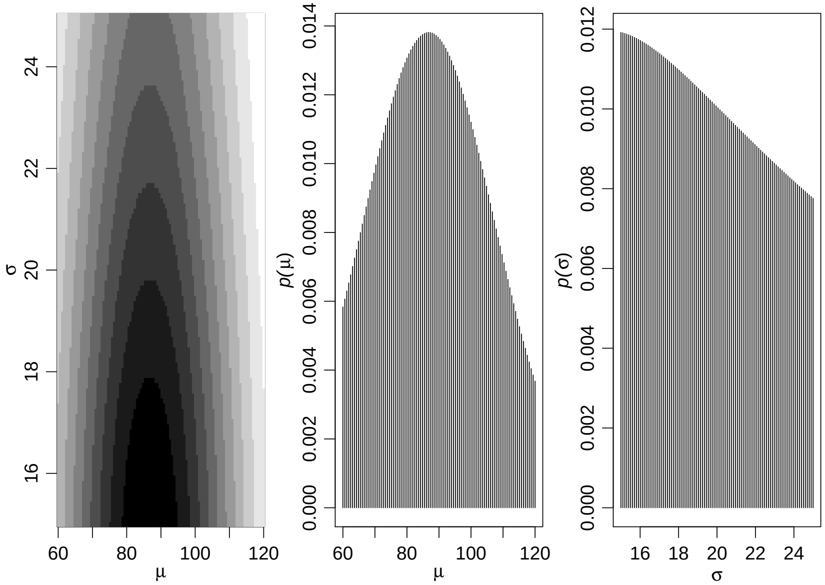
Joint and marginal prior distributions



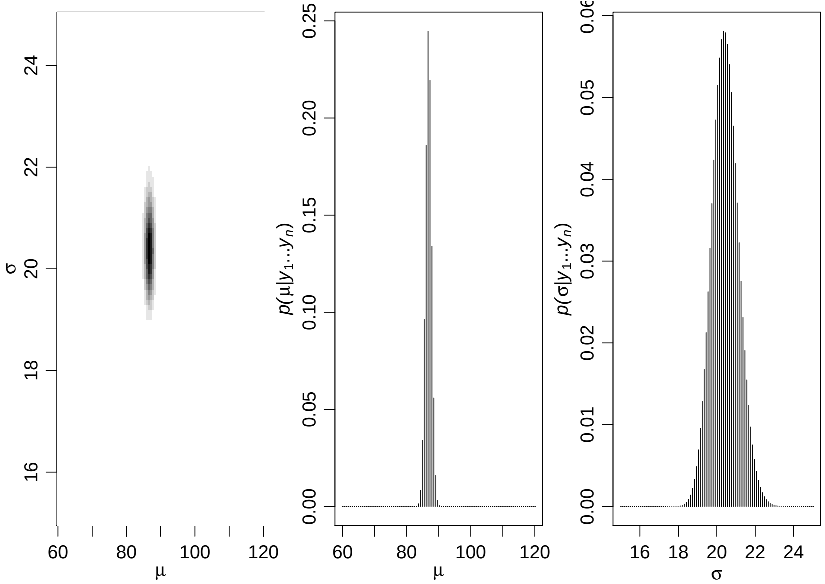
Joint and marginal posterior distributions



Finer grid: priors



Finer grid: posteriors



How did I get those previous graphs?

- Defined a grid of μ and σ values
  - e.g. first example for μ was `seq(60,120,length=10)`
- Calculated density at each grid point
  - e.g. using `dnorm`
- Standardized e.g. $p(\mu) = f(\mu)/\sum f(\mu)$
- Calculated prior grid $p(\mu) \cdot p(\sigma)$
- Calculated posterior grid $p(\mu, \sigma|y) = \frac{p(y|\mu,\sigma) \cdot p(\mu) \cdot p(\sigma)}{p(y)}$
- Calculated marginals of posterior grid by summing over relevant parameter e.g. $p(\mu|y) = \sum_\sigma p(\mu, \sigma|y)$

Now with continuous priors

$$p(\mu, \sigma | \mathbf{y}) = \frac{p(\mathbf{y}|\mu,\sigma)p(\mu,\sigma)}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\mu,\sigma)p(\mu,\sigma)}{\int_\mu \int_\sigma p(\mathbf{y}|\mu,\sigma)p(\mu,\sigma)d\sigma d\mu}$$

The bad news:

- Common choices of priors (e.g. what we have chosen) do not result in a closed-form expression for these posterior distributions

The good news:

- Not a problem if we can obtain a sample from the posterior distribution, which is very common in Bayesian inference.

Simulation based inference and Monte Carlo

The general idea in simulation-based inference: we can make inference about a random variable μ, using a sample $\mu^{(1)}, \ldots, \mu^{(S)}$ from its probability distribution. This is called a **Monte Carlo (MC)** approximation.

```
my_sample <- rnorm(5000, mean = 0, sd = 1)
mean(my_sample)
```

```
## [1] 0.01307762
```

```
sd(my_sample)
```

```
## [1] 1.012568
```

Monte Carlo

- Why can we use a sample mean as an approximation to the mean of a random variable?
- Just about any aspect of the distribution of μ can be approximated arbitrarily exactly with a large enough Monte Carlo sample, e.g.
  - the α-percentile of $\mu^{(1)}, \ldots, \mu^{(S)} \rightarrow$ the α-percentile of the distribution, e.g. the median
  - We can approximate $Pr(\mu \geq x)$ for any constant x by the proportion of samples for which $\mu \geq x$, because

$$1/S \sum_{s=1}^{S} I(\mu^{(s)} \geq x) \rightarrow Pr(\mu \geq x)$$

Monte Carlo

- With a simulation, it also becomes very easy to analyze the distributions of any function of 1 or more random variables, e.g.
  - use $1/\mu^{(s)}$ to study $1/\mu$
- Samples from marginal distributions may be obtained from samples from joint distributions, e.g.
  - the distribution of μ_1, where $(\mu_1, \mu_2) \sim N_2(\mathbf{0}, \Sigma)$ can be studied using samples of $\mu_1^{(s)}$ where $(\mu_1^{(s)}, \mu_2^{(s)}) \sim N_2(\mathbf{0}, \Sigma)$

Back to example

- Problem: For common choices of the priors on μ and σ, there is no closed-form expression for $p(\mu|y)$.
- Solution: let's obtain a posterior sample $\mu^{(1)}, \ldots, \mu^{(S)}$
- How to do this? Next week gives an overview.

Summary

Bayes rule for more than one parameter

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

- $p(\mathbf{y}) = \int_{\theta'} p(\mathbf{y}|\theta')p(\theta')d\theta'$
- The marginal posterior for just one parameter is given by $p(\theta_1|\mathbf{y}) = \int_{\theta_2'} \cdots \int_{\theta_p'} p(\theta_1, \theta_2', \ldots \theta_p'|\mathbf{y})d\theta_2' \ldots d\theta_p'$.
- Problem: often don't have a closed form solution for posterior
- Solution: We can make inference about any random variable θ, using a sample from its probability distribution. This is called a Monte Carlo (MC) approximation.
- If we are able to obtain a sample from posterior $p(\theta|\mathbf{y})$ then
  - for each parameter we have a sample of its marginal e.g. $\theta_1^{(1)}, \ldots, \theta_1^{(S)} \sim p(\theta_1|y)$.
  - we can report any summary we'd like, e.g. posterior mean (sample mean), posterior median or other percentiles (sample percentiles).