Bayesian Inference: Multiple Parameters

More Parameters

- Most statistical models have more than 1 parameter
  - can have thousands!
- We need to know how to deal with many parameters
  - actually a strength of the Bayesian approach
- It turns out to be simple in practice, but we need some maths first.

A Simple Example

- Estimating the mean and variance of a normal distribution
- Some data (heights in cm):
- 160, 170, 167, 162, 170, 171, 164, 175, 177, 178, 184, 174, 176, 186, 189, 197
- Fit a Normal distribution
  - 2 parameters, mean (μ) and variance (σ^2)
- Likelihood:

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

The Priors

- First, we will use a prior distribution that is uniform prior on μ and $\log(\sigma)$, for reasons that will become clear.
- The prior is thus

$$P(\mu,\sigma^2) \propto (\sigma^2)^{-1}$$

- This is not a proper probability distribution
  - the area under the curve is infinite!
- Can prove this still leads to a proper posterior
- Later we will use a different prior

The Posterior

- The posterior distribution for *n* observations is:

$$P(\mu,\sigma^2|y) \propto P(\mu,\sigma^2)P(y|\mu,\sigma^2)$$

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

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

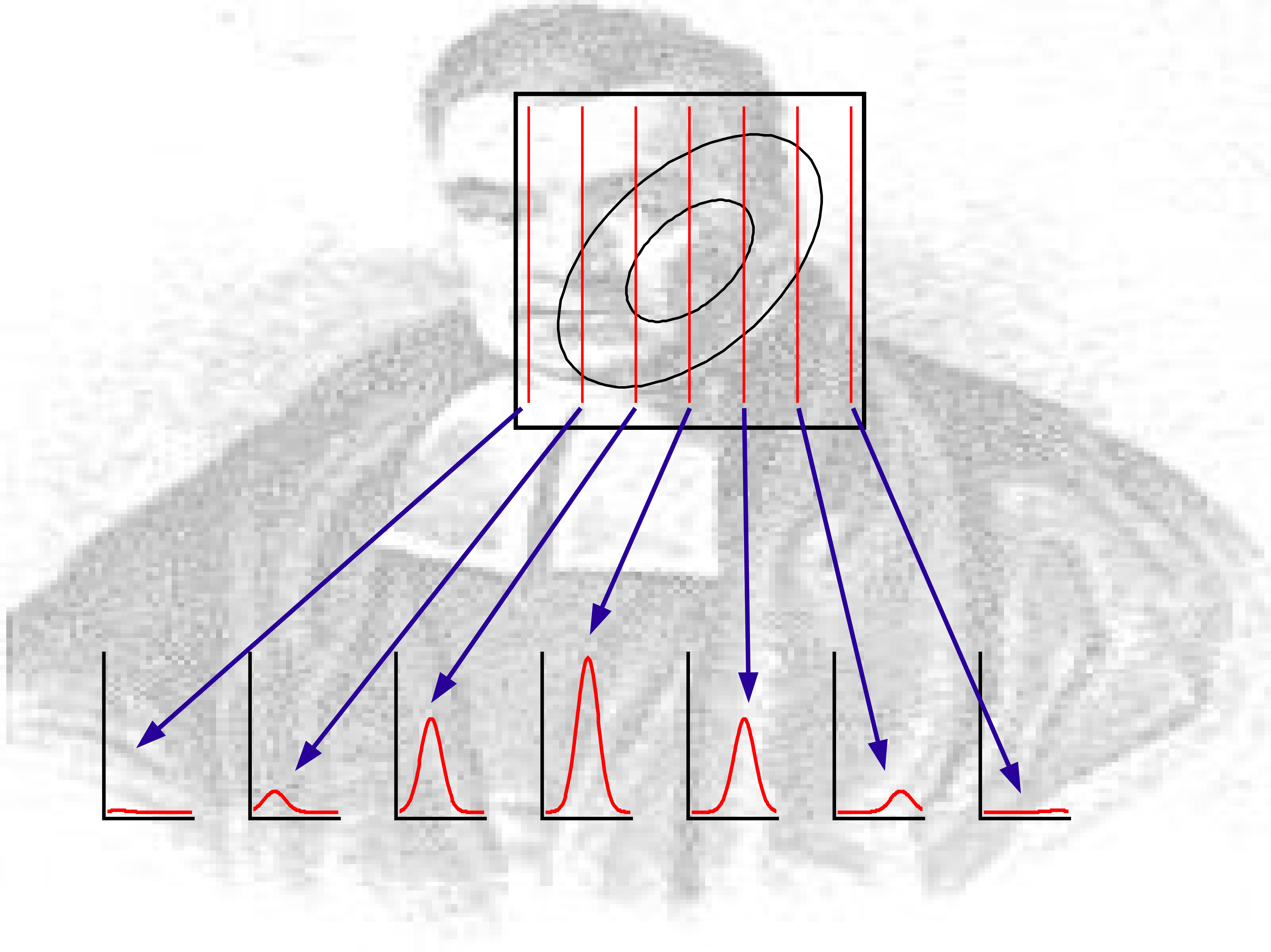
- This is the joint posterior for the parameters
- So, now we have an equation, what do we do with it?

Joint Distributions

- From the definition of a conditional distribution:

$$P(X_1, X_2) = P(X_1 | X_2) P(X_2)$$

- We can read this as slices of conditional distributions, weighted by the size of the slice

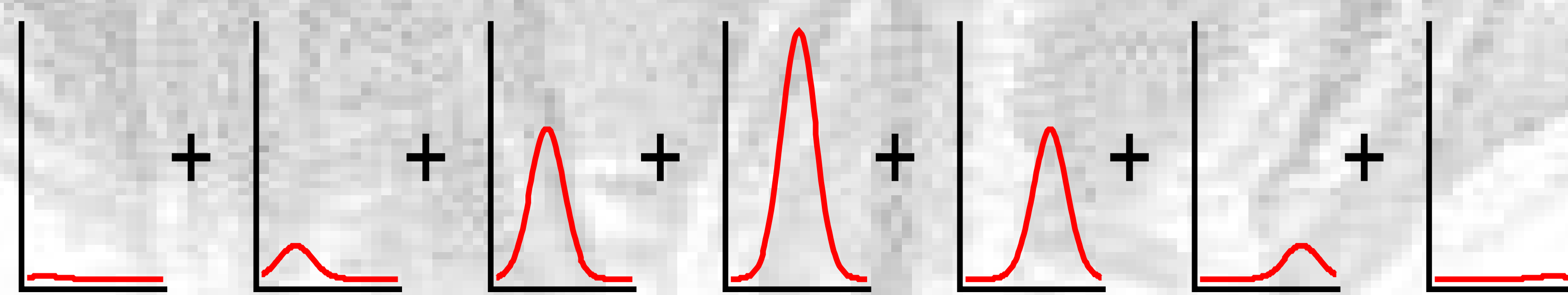


Marginal Distributions

- What if we want the distribution of one variable?
- We take the marginal distribution:

$$P(X_1) = \int P(X_1 | X_2) P(X_2) dX_2$$

- We just sum up the slices
  - but we need the marginal for X_2



The Normal Distribution

- We can get the marginal posteriors for μ and σ^2 by some maths:

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

$$= \frac{1}{\sigma^{n+2}} \exp\left(-\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{y} - \mu)^2 \right] \right)$$

- $\bar{y}$ and s^2 are the sample means and variances:

$$\bar{y} = \frac{\sum y_i}{n} \qquad s^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$$

Conditional Posterior for μ

- If we remove the constants that do not depend on μ then we get

$$P(\mu,|\sigma^2, y) \propto \exp\left(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right)$$

- Conveniently, this is just a Normal distribution. So

$$P(\mu,|\sigma^2 y) \sim N(\bar{y}, \sigma^2/n)$$

- Which still depends on σ^2

Marginal Posterior for σ^2

- For σ^2, we want to calculate

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

- This is easier than it looks. We get:

$$P(\sigma^2|y) \propto (\sigma)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2} \right)$$

- Which is an inverse gamma distribution
  - also known as a scaled inverse chi-squared

Marginal for μ

- Now we want

$$P(\mu|y) \propto \int P(\mu, \sigma^2|y)\, d\sigma^2$$

- Which, with some maths, becomes

$$P(\mu|y) \propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right]^{-n/2}$$

- This is just a *t* distribution, with *n*-1 df!

$$\frac{\mu - \bar{y}}{s/\sqrt{n}} \sim t_{n-1}$$

Prediction

- What if we want to predict a new value? We need

$$P(y_{new}|y) = \int\int P(y_{new}|\mu,\sigma^2)P(\mu,\sigma^2|y)\,d\sigma^2\,d\mu$$

- Which is also a t-distribution:

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

Summary

- We have the following posterior distributions:
- Conditional for μ
  - $\mu \sim N(\bar{y}, \sigma^2)$
- Marginal for μ
  - $\mu \sim t_{n-1}(\bar{y}, s^2/n)$
- Marginal for σ^2
  - $\sigma^2 \sim \chi^2\,(n\text{-}1, s^2)$
- Predictive
  - $y_{new} \sim t_{n-1}(\bar{y}, (1+1/n)s^2)$

Simulation

- It is often easier to deal with these distributions by simulation
  - sometimes easier to see what is going on
- Here we need RNGs for normal and chi-squared distributions
- Most real work is done by simulation
- Draw the parameters many times from the right distribution
- The easy one: Conditional for $\mu_s \mid \sigma^2$
  - Simulate from a $N(\bar{y}, \sigma^2)$

Simulation

- Marginal for μ

1. Draw σ_s^{-2} from a χ^2 $(n\text{-}1, s^2)$
2. Draw μ_s from a N$(\bar{y}, \sigma_s^{-2})$

- Marginal for σ_s^2
  - Draw σ_s^{-2} from a $\chi^2(n\text{-}1, s^2)$, then take inverse
- Predictive for y_{new}

1. Draw σ_s^{-2} and μ_s as above
2. Draw y_{new} from a N(μ_s, σ_s^{-2})

Different Priors

- What if we want informative priors
  - e.g. if we have information to use?
- One possible set of priors:
  - $\mu|\sigma^2 \sim N(\mu_0, \sigma_0^2/\kappa_0)$
  - $\sigma^2 \sim \text{Inv-}\chi^2(\upsilon_0, \sigma_0^2)$
- Why these priors? Because the posterior has the same distribution
  - conjugate

Different Posteriors

- $\mu|y,\sigma^2 \sim \mathrm{N}(\mu_n, \sigma^2/\kappa_n)$

$$\mu_n = \frac{\kappa_0 \mu_0 + n\bar{y}}{\kappa_0 + n} \qquad \kappa_n = \kappa_0 + n$$

- $\sigma^2|y \sim \mathrm{Inv}\text{-}\chi^2(\nu_n, \sigma_n^2)$

$$\nu_n = \nu_0 + n \qquad \nu_n \sigma_n^2 = \nu_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2$$

- $\mu|y \sim t_{\nu_n}(\mu_n, \sigma_n^2/\kappa_n)$

What do we want?

- The basic calculations are as outlined above
- From the joint posterior, we calculate the marginal distributions
  - can calculate a joint distribution for a subset of the parameters by marginalising over the rest
- For real models the calculations get difficult
- Instead, we use simulation
  - makes things easier

Marginalisation by Simulation

- To simulate a conditional distribution, we plug in the parameters we are conditioning on:
- Simulate $P(\mu|\sigma^2, y) \sim N(\bar{y}, \sigma^2/n)$ by plugging in $\bar{y}$ and σ^2/n
- To simulate the marginal, we use the relationship

$$P(X_1) = \int P(X_1|X_2) P(X_2) dX_2$$

- If we can do $P(X_1|X_2)$ and $P(X_2)$, then we just draw X_2, then $X_1|X_2$ and repeat this many times
  - we just ignore X_2