NORMAL AND MULTIVARIATE MODELS

BST228 Applied Bayesian Analysis

Speaker notes

No notes on this slide.

RECAP

- Finalized binomial and Poisson models, including prior sensitivity.
- Derived normal likelihood for i.i.d. observations.
- Posterior for location parameter μ given data $\mathbf{y}$ and known precision τ (almost!).

Speaker notes

No notes on this slide.

OUTLINE

- Method for estimating hyperparameters for weakly informative priors.
- Posterior for location parameter μ given data and known precision τ.
- Posterior for precision parameter τ given data and known location μ.
- Joint inference and marginal distributions.

Speaker notes

No notes on this slide.

```
> # Declare the target quantiles and values.
> limits <- c(1e-6, 1e-3)
> qs <- c(0.025, 0.975)
> # Find and report parameters.
> result <- optim(
+   c(0, 0),
+   function(par) {
+     par <- exp(par)
+     residuals <- pbeta(limits, par[1], par[2]) -
+       qs
+     sum(residuals^2)
+   },
+ )
> exp(result$par)
[1]    0.6487655 2897.5668473
>
```

Speaker notes

- Lines #2-3 declare the quantiles of the prior and quantile values.
- The `optim` function performs the optimization starting at the vector of zeros in #6.
- #7-11 evaluate the mean squared error between targets `qs` and actual CDF values evaluated at `limits`. #8 applies an `exp` transform to ensure parameters of the beta distribution are positive. This is required because `optim` operates without constraints.
- #15 reports the optimized values, applying `exp` for consistency with the optimization.

POSTERIOR FOR μ UNDER NORMAL LIKELIHOOD WITH KNOWN τ

The posterior given a normal prior $p\left(\mu \mid \nu_0, \kappa_0\right)$ with prior mean ν_0 and precision κ_0 is

$$p\left(\mu \mid \mathbf{y}, \tau\right) \propto \exp\left(-\frac{\kappa_0}{2}\left(\mu^2 - 2\mu\nu_0\right)\right) \exp\left(-\frac{n\tau}{2}\left(\mu^2 - 2\mu\bar{y}\right)\right),$$

where we have expanded the square in the exponential of the prior. Combining the exponentials and collecting terms in μ and μ^2 yields

$$\begin{aligned} p\left(\mu \mid \mathbf{y}, \tau\right) &\propto \exp\left(-\frac{(\kappa_0 + n\tau)\mu^2 - 2\mu(\kappa_0\nu_0 + n\tau\bar{y})}{2}\right) \\ &\propto \exp\left(-\frac{\kappa_0 + n\tau}{2}\left(\mu^2 - 2\mu\frac{\kappa_0\nu_0 + n\tau\bar{y}}{n\tau + \kappa_0}\right)\right). \end{aligned}$$

Comparing with the functional form of a normal distribution, we find that the posterior has mean $\nu_n = \frac{\kappa_0\nu_0 + n\tau\bar{y}}{n\tau + \kappa_0}$ and precision $\kappa_n = \kappa_0 + n\tau$.

Speaker notes

No notes on this slide.

Update rules for parameters of posterior on μ given known precision τ are

$$\nu_n = \frac{\kappa_0 \nu_0 + n\tau \bar{y}}{n\tau + \kappa_0},$$

$$\kappa_n = \kappa_0 + n\tau.$$

Speaker notes

- The posterior mean ν_n is the average of the prior mean ν_0 and sample mean $\bar{y}$ weighted by the prior and likelihood precisions.
- The more data we observe (increasing n) or the more precise the observations (increasing τ), the closer the posterior mean is to the sample mean.
- For large n, the posterior variance $\kappa_n^{-1} \propto n^{-1}$, and we recover the familiar square-root scaling of the standard error.

Update rules with likelihood scale σ and prior scale ρ_0 are

$$\nu_n = \frac{\frac{n\bar{y}}{\sigma^2} + \frac{\nu_0}{\rho_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\rho_0^2}},$$

$$\rho_n = \frac{1}{\sqrt{\frac{1}{\rho_0^2} + \frac{n}{\sigma^2}}},$$

much more tedious than using precision instead of scale.

Speaker notes

- I claimed without evidence that parameterizing the normal distribution using a precision instead of scale parameter was nicer.
- The parameter update rules using prior scale $\rho_0 = \kappa_0^{-1/2}$ and likelihood scale $\sigma = \tau^{-1/2}$ are messier and less interpretable.

PAIRED EXERCISE

An instrument with known measurement noise $\sigma = 0.1$ yields replicates $\mathbf{y} = (2.1, 2.5, 1.6, 1.7)$ for a chemical concentration.

- What is the posterior mean and standard deviation?
- What hyperparameters did you use?
- What issues might arise due to using a normal model?
- How significant are these issues? How would your answer change for $\mathbf{y} = (0.1, 0.2, 0.1)$.

Speaker notes

No notes on this slide.

```
> # Declare data and known noise level.
> y <- c(2.1, 2.5, 1.6, 1.7)
> sigma <- 0.1
> # Define hyperparameters.
> nu_0 <- 0
> kappa_0 <- 1e-4
> # Update parameters and sample.
> n <- length(y)
> y_bar <- mean(y)
> tau <- 1 / sigma ^ 2
> kappa_n <- kappa_0 + n * tau
> nu_n <- (n * tau * y_bar + kappa_0 * nu_0) /
+   (kappa_0 + n * tau)
> mu_samples <- rnorm(1000, nu_n, 1 / sqrt(kappa_n))
> # Report summary statistics.
> c(mean(mu_samples), sd(mu_samples))
[1] 1.97701148 0.05082841
>
```

Speaker notes

- Lines #2-3 declare the data; #5-6 declare hyperparameters.
- #8-13 evaluate the posterior parameters.
- #14 draws posterior samples, and #16 reports the posterior mean.

RECAP

- Normal likelihood is one of the most commonly used distributions.
- For known τ, we derived a conjugate prior and parameter update rules for μ.
- We combined evidence from multiple independent observations.

Speaker notes

No notes on this slide.

POSTERIOR FOR τ UNDER NORMAL LIKELIHOOD WITH KNOWN μ

Consider the posterior

$$\begin{aligned} p\left(\tau \mid \mathbf{y}, \mu\right) &\propto p\left(\tau\right) \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left(-\frac{\tau}{2} \sum_{i=1}^{n} \left(y_i - \mu\right)^2\right) \\ &\propto p\left(\tau\right) \tau^{n/2} \exp\left(-\frac{\tau S}{2}\right), \end{aligned}$$

where $S = \sum_{i=1}^{n} \left(y_i - \mu\right)^2$. The kernel matches a gamma distribution, and we use a gamma prior $\tau \sim \textsf{Gamma}\left(a_0, b_0\right)$. Collecting terms yields

$$p\left(\tau \mid \mathbf{y}, \mu\right) \propto \tau^{a_0 + n/2 - 1} \exp\left(-\tau \left(b_0 + \frac{S}{2}\right)\right),$$

and the posterior is $\theta \mid \mathbf{y}, \mu \sim \textsf{Gamma}\left(a_0 + \frac{n}{2}, b_0 + \frac{S}{2}\right)$.

Speaker notes

No notes on this slide.

Update rules for τ posterior parameters given known mean μ are

$$a_n = a_0 + \frac{n}{2},$$

$$b_n = b_0 + \frac{1}{2}\sum_{i=1}^{n}\left(y_i - \mu\right)^2.$$

Speaker notes

- We can verify that, in the large sample limit, the posterior mean is $\frac{a_n}{b_n} = \left[\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \mu\right)^2\right]^{-1}$, i.e., the inverse sample variance.

PAIRED EXERCISE

A sample has known chemical concentration $\mu = 2$ and measurements are as before, i.e., $\mathbf{y} = (2.1, 2.5, 1.6, 1.7)$.

- What is the posterior mean and variance?
- What hyperparameters did you use?
- Is the inferred precision consistent with $\sigma = 0.1$ from before?

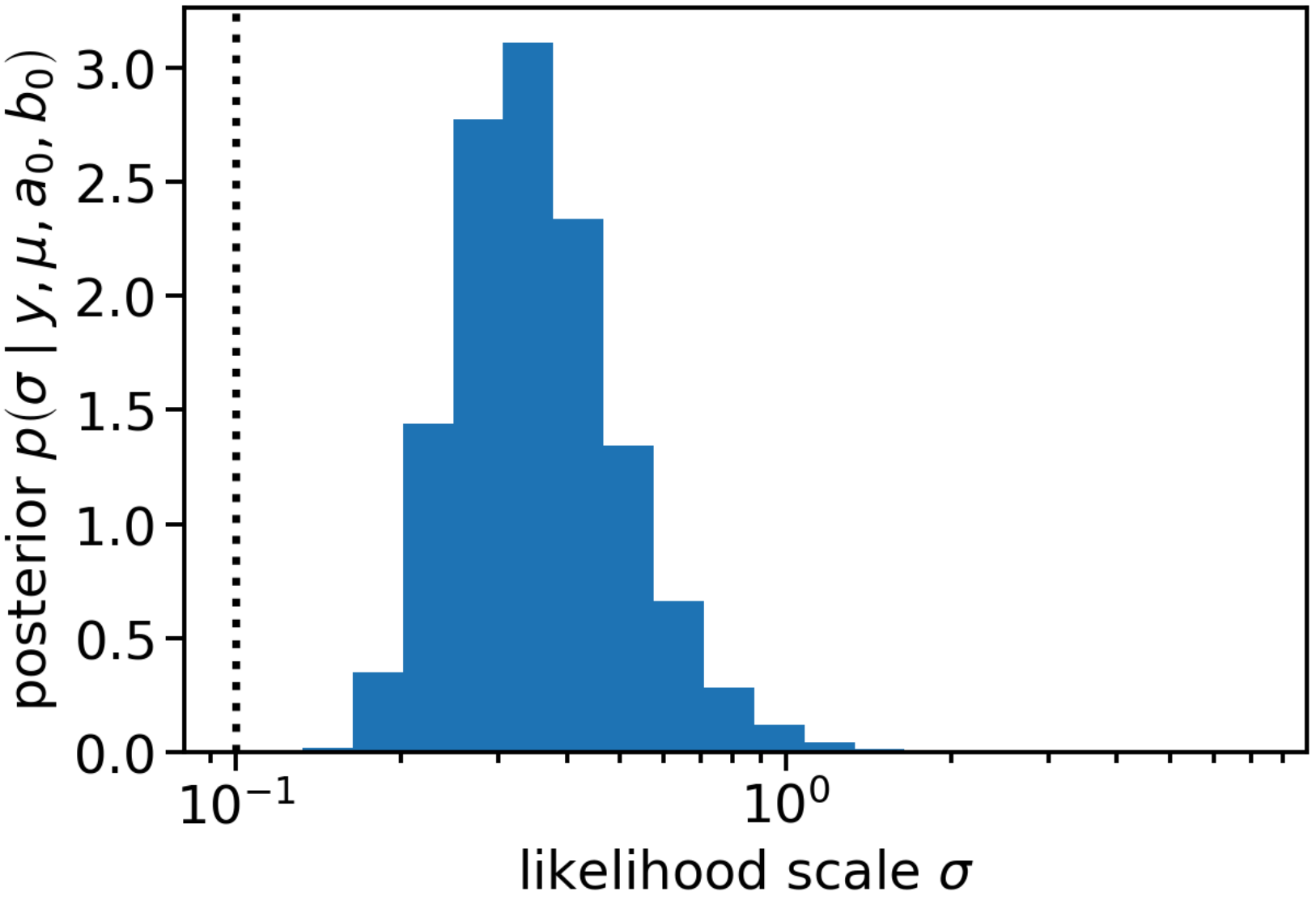
Speaker notes

No notes on this slide.

```
> # Declare data.
> y <- c(2.1, 2.5, 1.6, 1.7)
> n <- length(y)
> mu <- 2
> # Hyperparameters.
> a_0 <- 0.001
> b_0 <- 0.001
> # Update and sample
> a_n <- a_0 + n / 2
> b_n <- b_0 + sum((y - mu) ^ 2) / 2
> tau_samples <- rgamma(1000, a_n, b_n)
> c(mean(tau_samples), sd(tau_samples))
[1] 8.005844 5.591596
> # Compare with reported noise level.
> sigma_samples <- 1 / sqrt(tau_samples)
> c(mean(sigma_samples), sd(sigma_samples),
+   mean(sigma_samples < 0.1))
[1] 0.440847 0.221100 0.000000
>
```

Speaker notes

- Lines #2-4 declare data and known mean.
- #6-7 declare hyperparameters.
- #9-12 evaluate posterior parameters, sample, and report posterior mean.
- #15 transforms precision samples to scale samples.
- #16 reports posterior mean, standard deviation, and the fraction of scale samples that are smaller than the noise scale reported by the instrument manufacturer.
- All posterior samples of σ exceed the reported noise scale, suggesting that the manufacturer overstated the precision of their instrument.



Speaker notes

- The tension between reported and inferred observation noise scale is obvious from a histogram of posterior samples.
- This exemplifies why we often want to infer parameter jointly, i.e., location parameter μ and precision τ at the same time.

RECAP

- Conjugate prior for precision of normal likelihood with known mean is gamma.
- Inverse gamma on the variance is an alternative prior.

Speaker notes

No notes on this slide.