Data illustration of Bayesian inference

Consider the acid concentration data example discussed in Examples 8.2.3, 8.5.4 and 8.6.2 of DeGroot and Schervish (2012). The data set we work with here combines the 10 measurements from Example 8.5.4 with the 20 measurements given in Exercise 8.6.16 for the full sample of size \( n = 30 \).

To illustrate Bayesian analysis for a normal distribution under the conjugate prior, we assume a \( \mathcal{N}(\mu, \tau) \) distribution for lactic acid concentration, where \( \mu \) is the mean and \( \tau \) is the precision parameter (hence, the variance \( \sigma^2 = 1/\tau \)). Simple exploratory analysis of the data does not indicate obvious deviations from symmetry and unimodality for the underlying distribution, and thus the normal distribution assumption provides a reasonable starting point.

Prior specification. The Bayesian model is applied with a \( \text{gamma}(\alpha_0, \beta_0) \) prior for \( \tau \) and a \( \mathcal{N}(\mu_0, \lambda_0 \tau) \) conditional prior for \( \mu \) given \( \tau \), where \( \alpha_0 = 2 \), \( \beta_0 = 0.125 \), \( \mu_0 = 1.45 \) and \( \lambda_0 = 4.5 \). Details for this choice of prior hyperparameters will be given in class. Note that the value of \( \alpha_0 = 2 \) ensures finite marginal mean and variance for \( \mu \), but yields infinite prior variance for \( \sigma^2 \); recall that the \( \text{gamma}(\alpha_0, \beta_0) \) prior for \( \tau \) implies an \( \text{inverse-gamma}(\alpha_0, \beta_0) \) prior for \( \sigma^2 \). The remaining prior hyperparameters are specified based on a “prior guess” of \((0.7, 2.2)\) for the range of plausible lactic acid concentration values; in the absence of actual prior information in this illustrative example, the interval above is chosen by expanding the range of observed values in the data. A useful means to study the combined effect of the priors for the model parameters is through the implied prior predictive distribution; refer to Figure 2 and the details given in the relevant section below.

Inferences based on the posterior distribution. Under the specific prior given above, the posterior distribution for \((\mu, \tau)\) is analytically available through

\[
\xi(\mu, \tau \mid x) = \mathcal{N}(\mu \mid 1.443, 34.5\tau) \text{gamma}(\tau \mid 17, 1.46)
\]

and we can also obtain the marginal posterior distribution, \( \xi(\mu \mid x) \), from the transformation \( 20.04(\mu - 1.443) \mid x \sim t_{34} \) (refer to Theorems 8.6.1 and 8.6.2). The joint posterior density, \( \xi(\mu, \tau \mid x) \), and marginal posterior densities, \( \xi(\mu \mid x) \) and \( \xi(\tau \mid x) \), can be plotted and compared with the respective prior densities. Point and interval estimates for the parameters can be computed using moments and percentiles of the \( t \) and gamma marginal posterior distributions. For instance, the 0.975 percentile of the \( t_{34} \) distribution, \( T_{34}^{-1}(0.975) \approx 2.033 \), and therefore \( \Pr(-2.033 < 20.04(\mu - 1.443) < 2.033 \mid x) = 0.95 \), from which we obtain \( \Pr(1.3416 < \mu < 1.5444 \mid x) = 0.95 \). Although the interpretation is different, the endpoints of the posterior probability interval can be contrasted with the 95% confidence interval for
Figure 1: Simulation-based inference for the model parameters: bivariate plot of the posterior samples for \((\mu, \tau)\) (upper left panel), probability histogram of samples from \(\xi(\mu | x)\) (upper right panel), from \(\xi(\tau | x)\) (lower left panel), and from \(\xi(\sigma | x)\) (lower right panel). The red line in the upper right, lower left and lower right panels indicates the respective prior density.

\(\mu\) (based on the \(t_{20}\) distribution) given by \((1.3287, 1.5553)\). The results become even more similar as the prior distribution for \((\mu, \tau)\) becomes more dispersed, that is, as the amount of prior information is reduced.

**Sampling from the posterior distribution.** Inference for the model parameters and posterior predictive inference are facilitated by *posterior simulation*, that is, by sampling from the posterior distribution, \(\xi(\mu, \tau | x)\), and using the posterior samples for simulation-based inference and prediction. In the context of the specific problem, direct sampling from \(\xi(\mu, \tau | x)\) is straightforward using its decomposition in terms of the conditional posterior distribution for \(\mu\) given \(\tau\) and the marginal posterior distribution for \(\tau\). In particular, to collect \(B\) posterior samples, \(\{ (\mu_b, \tau_b) : b = 1, ..., B \} \): for \(b = 1, ..., B\), we draw \(\tau_b \sim \text{gamma}(17, 1.46)\); and then draw \(\mu_b \sim \text{N}(1.443, 34.5\tau_b)\). Note that \(\{ \mu_b : b = 1, ..., B \}\) is (by definition) a sample from the marginal posterior distribution \(\xi(\mu | x)\), and we can therefore obtain inference for the mean even if we do not have access to the \(t\) distribution result. In particular, using the empirical 0.025 and 0.975 percentiles from the posterior samples for \(\mu\), we obtain
Figure 2: Histogram of the lactic acid concentration data (left panel) along with simulation-based estimates of the prior and posterior predictive density (middle and right panels, respectively).

\[ \Pr(1.3409 < \mu < 1.5432 \mid \mathbf{x}) = 0.95, \] which is essentially the same interval with the one based on the analytically available distribution for \( \xi(\mu \mid \mathbf{x}) \). Moreover, the posterior density of any function of the parameters can be estimated by simply evaluating the function on the posterior samples; see Figure 1 for an illustration with the standard deviation, \( \sigma = \tau^{-1/2} \).

**Prior and posterior predictive distributions.** The posterior samples for \((\mu, \tau)\) can be used to sample from the posterior predictive distribution. Recall the form of the posterior predictive density:

\[ f(x_0 \mid \mathbf{x}) = \int \int N(x_0 \mid \mu, \tau) \xi(\mu, \tau \mid \mathbf{x}) \, d\mu \, d\tau. \]

Therefore, if for every sample \((\mu_b, \tau_b)\) from \( \xi(\mu, \tau \mid \mathbf{x}) \) we draw \( x_{0,b} \) from \( N(\mu_b, \tau_b) \), the resulting \( \{x_{0,b} : b = 1, \ldots, B\} \) provide a sample from the posterior predictive distribution which can be used for predictive inference given the data. The right panel of Figure 2 plots the simulation-based estimate of the posterior predictive density in the form of a histogram based on the posterior predictive samples. Following a similar approach where the posterior samples are replaced with samples from the prior distribution, \( \xi(\mu, \tau) \), we can estimate the prior predictive density, \( f(x_0) = \int \int N(x_0 \mid \mu, \tau) \xi(\mu, \tau) \, d\mu \, d\tau \); see the middle panel of Figure 2 for the results under the prior hyperparameters discussed above.
R code

R code to estimate the prior predictive density:

```r
> B <- 10000
> pr.mu <- rep(0,B)
> pr.tau <- rep(0,B)
> pr.pred <- rep(0,B)
> for(i in 1:B)
+ {
+   pr.tau[i] <- rgamma(1,shape=2,rate=0.125)
+   pr.mu[i] <- rnorm(1,mean=1.45,sd=sqrt(1/(4.5*pr.tau[i])))
+   pr.pred[i] <- rnorm(1,mean=pr.mu[i],sd=sqrt(1/pr.tau[i]))
+ }
> hist(prob=T,main="",xlab="Prior predictive density",xlim=c(0,3),ylim=c(0,1.3),pr.pred)
```

and for the posterior predictive density:

```r
> B <- 10000
> mu <- rep(0,B)
> tau <- rep(0,B)
> post.pred <- rep(0,B)
> for(i in 1:B)
+ {
+   tau[i] <- rgamma(1,shape=17,rate=1.46)
+   mu[i] <- rnorm(1,mean=1.443,sd=sqrt(1/(34.5*tau[i])))
+   post.pred[i] <- rnorm(1,mean=mu[i],sd=sqrt(1/tau[i]))
+ }
> hist(prob=T,main="",xlab="Posterior predictive density",xlim=c(0,3),ylim=c(0,1.3),post.pred)
```