Chapter 8.6 Bayesian Inferences for a Mean

Jim Albert and Monika Hu

Chapter 8 Modeling Measurement and Count Data

Back to our tennis example

- Our Normal prior had mean 18 seconds and standard deviation 1.56 seconds.
- After collecting 20 time-to-serve measurements with a sample mean of 17.2, the posterior distribution $\mathrm{Normal}(17.4, 0.77)$ reflects our opinion about the mean time-to-serve.

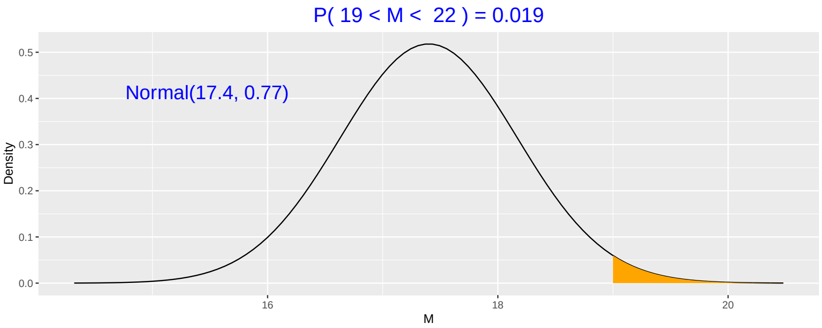
Bayesian inferences and prediction

- Bayesian inferences about the mean μ are based on various summaries of this posterior Normal distribution.
- It is convenient to use R functions such as `pnorm()` and `qnorm()` to conduct Bayesian hypothesis testing and construct Bayesian credible intervals.
- Simulation-based methods utilizing functions such as `rnorm()` are also useful to provide approximations to those inferences.
- Predictions of future data are also of interest. For example, one might want to predict the next time-to-serve measurement

Bayesian hypothesis testing

- In a *testing* problem, one wishes to check the validity of a statement about a population quantity. Someone says Federer takes on average at least 19 seconds to serve. Is this reasonable?
- The current beliefs about Federer's mean time-to-serve are summarized by a Normal(17.4, 0.77) distribution.
- To assess if "μ is 19 seconds or more" is reasonable, one computes its probability of $\mu \geq 19$ under the posterior curve.

Computation of posterior probability



- This probability is about 0.019, a small value, so one would conclude that this person's statement is unlikely to be true.

Simulation approach

- Simulation provides an alternative approach to obtaining the probability $Prob(\mu \geq 19 \mid \mu_n = 17.4, \sigma_n = 0.77)$.
- One generates 1000 values from the $\textrm{Normal}(17.4, 0.77)$ distribution and approximates the probability of "μ is 19 seconds or more" by computing the percentage of values that falls above 19.

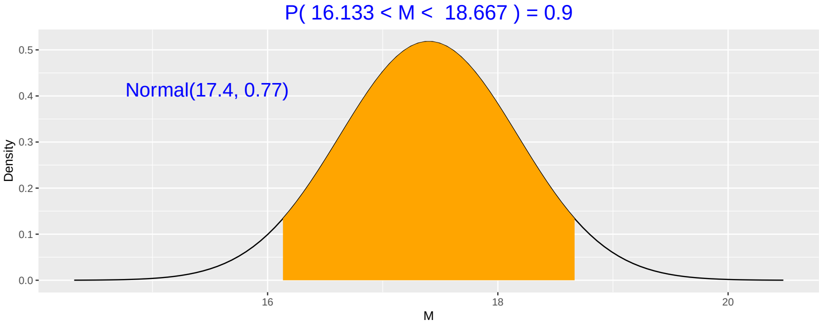
```
S <- 1000
NormalSamples <- rnorm(S, 17.4, 0.77)
sum(NormalSamples >= 19) / S
```

```
## [1] 0.024
```

- The value of 0.024 is a simulation approximation to the exact answer of 0.019 using the `pnorm()` function.

A Bayesian interval estimate

- The `normal_interval()` function can be used to show an interval that contains μ with a specific probability.



Using simulation

- For simulation-based inference, one generates a large number of values from its posterior distribution, then finds the 5th and 95th sample quantiles to obtain the middle 90% of the generated values.

```
S <- 1000
NormalSamples <- rnorm(S, 17.4, 0.77)
quantile(NormalSamples, c(0.05, 0.95))
```

```
##        5%      95%
## 16.15061 18.69062
```

Credible intervals to test hypotheses

- Suppose one again wants to evaluate the statement "Federer takes on average at least 19 seconds to serve."
- One computes the 90% credible interval and notes that the values of μ "at least 19" are not included in the exact 90% credible interval (16.15, 18.69).

```
qnorm(c(0.025, 0.975), 17.4, 0.77)
```

```
## [1] 15.89083 18.90917
```

- On the basis of this credible interval calculation, one concludes that the statement about Federer's time-to-serve is unlikely to be true.

Prediction

- Suppose one is interested in predicting Federer's future time-to-serve.
- Since one has already updated the belief about the parameter μ, the prediction is made based on its posterior predictive distribution.
- One approach derives the exact posterior predictive distribution $f(\tilde{Y} = \tilde{y} \mid Y = y)$.
- The second approach is a simulation-based approach.

Exact predictive distribution

- Consider making a prediction of a single Federer's time-to-serve $\tilde{Y}$.
- Suppose the sampling density of $\tilde{Y}$ given μ and σ is $f(\tilde{Y} = \tilde{y} \mid \mu)$ and suppose the current beliefs about μ are represented by the density $\pi(\mu)$.
- One can compute the joint density of $(\tilde{y}, \mu)$ is finding the product $f(\tilde{y}, \mu) = f(\tilde{Y} = \tilde{y} \mid \mu)\pi(\mu)$ and integrating out μ
- One finds that the predictive density for $\tilde{Y}$ is Normal with mean and standard deviation given by

$$E(\tilde{Y}) = \mu_0, \; SD(\tilde{Y}) = \sqrt{\sigma^2 + \sigma_0^2}.$$

Posterior predictive density

- After observing the sample values $y_1, \cdots, y_n$, the current beliefs about the mean μ are represented by a Normal(μ_n, σ_n) density, where the mean and standard deviation are given by

$$\mu_n = \frac{\phi_0 \mu_0 + n\phi\bar{y}}{\phi_0 + n\phi}, \sigma_n = \sqrt{\frac{1}{\phi_0 + n\phi}}.$$

- Then the posterior predictive density of the single future observation $\tilde{Y}$ is Normal with mean μ_n and standard deviation $\sqrt{\sigma^2 + \sigma_n^2}$. That is,

$$\tilde{Y} = \tilde{y} \mid y_1, \cdots, y_n, \sigma \sim \mathrm{Normal}(\mu_n, \sqrt{\sigma^2 + \sigma_n^2}).$$

Two sources of uncertainty

- The variance of the future $\tilde{Y}$ is given by $\sigma^2 + \sigma_n^2$.
- There are two sources of variability represented: (1) the data model variance σ^2, and (2) the posterior variance σ_n^2.
- If one allows the sample size n to grow, the posterior variance will approach zero.
- In this "large n" case, the uncertainty in inference about the population mean μ will decrease; however the uncertainty in prediction will approach the sampling variance σ^2.

Predictions by simulation

- An alternative method of computing the predictive distribution is by simulation.
- One can simulates a value from the posterior predictive distribution in two steps:

1. Sample a value of μ from its posterior distribution

$$\mu \sim \mathrm{Normal}\left(\frac{\phi_0\mu_0 + n\phi\bar{y}}{\phi_0 + n\phi}, \sqrt{\frac{1}{\phi_0 + n\phi}}\right),$$

2. Sample a new observation $\tilde{Y}$ from the data model (i.e. a prediction)

$$\tilde{Y} \sim \mathrm{Normal}(\mu, \sigma).$$

Using R

- This two-step procedure for simulating 1000 predictions is implemented for our time-to-serve example using the following R script.

```
S <- 1000
pred_mu_sim <- rnorm(S, mu_n, sigma_n)
pred_y_sim <- rnorm(S, pred_mu_sim, sigma)
```

The vector `pred_y_sim` contains 1000 predictions of Federer's time-to-serve.

Comparison of exact predictive density and density estimate of simulated predictions

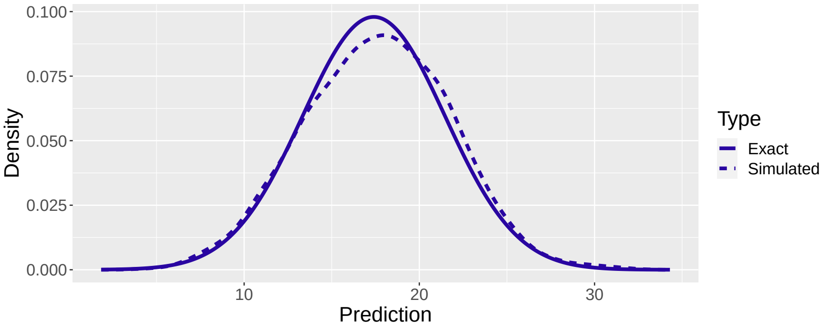


Figure 1: Display of the exact and simulated time-to-serve for Federer's example.