Imperial College London

Problem Sheet 4 Solutions

MATH50011
Statistical Modelling 1

Week 4

Lecture 7 (Proof of MLE Consistency and Asymptotic Normality)

1. In the lecture notes, we saw that MLEs are asymptotically normal and sketched a proof of this (subject to regularity conditions). Many other estimators are also the solutions to estimating equations.

   Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables and suppose that we wish to estimate the value of $\theta_0 \in \mathbb{R}$ defined as the unique $E[\psi(X_1, \theta)] = 0$ for a twice continuously differentiable function $\psi : \mathbb{R}^2 \to \mathbb{R}$. Define $\hat{\theta}_n$ as the unique solution to $\sum_{i=1}^n \psi(X_i, \theta) = 0$.

   Sketch a proof that $\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N(0, \sigma^2(\theta_0))$ and find an expression for $\sigma^2(\theta_0)$. At what steps in your proof sketch do you need additional assumptions required to justify an operation?

   Solution. We largely follow the steps for the proof of asymptotic normality of the MLE. We begin with a first-order Taylor expansion/mean value theorem so that

   $$0 = \sum_{i=1}^n \psi(X_i, \hat{\theta}_n) = \sum_{i=1}^n \psi(X_i, \theta_0) + \sum_{i=1}^n \frac{\partial}{\partial \theta} \psi(X_i, \theta)\Big|_{\theta=\tilde{\theta}_n} (\hat{\theta}_n - \theta_0)$$

   for some $\tilde{\theta}_n$ between $\hat{\theta}_n$ and θ_0. Let

   $$\dot{\psi}(\tilde{\theta}_n) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \psi(X_i, \theta)\Big|_{\theta=\tilde{\theta}_n}.$$

   After some rearranging, we have that

   $$-\dot{\psi}(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0) = \frac{1}{n} \sum_{i=1}^n \psi(X_i, \theta_0).$$

   The RHS above has mean zero by assumption. Provided

   $$Var\{\psi(X_i, \theta_0)\} = E\{\psi(X_i, \theta_0)^2\} = A(\theta_0)$$

   exists and is finite, we can apply the central limit theorem to find

   $$-\dot{\psi}(\tilde{\theta}_n)\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(X_i, \theta_0) \to_d N(0, A(\theta_0)).$$

If we have that $-\dot{\Psi}(\tilde{\theta}_n) \to_p B(\theta_0) = E\left\{ \frac{\partial}{\partial\theta}\psi(X_i,\theta)\Big|_{\theta=\theta_0} \right\}$ where $0 < B(\theta_0) < \infty$, then by Slutsky's lemma

$$\frac{-\dot{\Psi}(\tilde{\theta}_n)}{B(\theta_0)} B(\theta_0)\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N(0, A(\theta_0)).$$

Hence, we also have

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d N(0, B^{-2}(\theta_0)A(\theta_0))$$

so that $\sigma^2(\theta_0) = B^{-2}(\theta_0)A(\theta_0)$.

Note that:

- A 2nd-order Taylor expansion could also be used at the beginning of the proof, which then requires that the higher-order term converges to zero in probability.
- The assumption that $-\dot{\Psi}(\tilde{\theta}_n)$ converges in probability could be stated in similar equivalent ways (such as we did for the second partial derivatives of the log-likelihood in the lecture notes).
- For a regular parametric model, with $\psi(x;\theta) = \frac{\partial}{\partial\theta}\log f(x;\theta)$ we would have $A(\theta_0) = B(\theta_0) = I_f(\theta_0)$ so that $\sigma^2(\theta_0) = I_f(\theta_0)^{-1}$ as expected.

2. In the notation of Problem Sheet 3.10 (the R lab), define the one-step estimator

$$\hat{\theta}_n^{(1)} = T_n + I_n(T_n)^{-1}U_n(T_n).$$

Suppose that T_n is asymptotically normal and that $I_n(T_n)$ is consistent for the Fisher information $I_f(\theta)$. Show that

$$\sqrt{n}(\hat{\theta}_n^{(1)} - \theta_0) \to_d N(0, I_f(\theta_0)^{-1}).$$

Hint: use a first-order Taylor expansion of $U_n(\theta)$.

Solution. Recall that

$$U_n(\theta) = \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f_\theta(X_i)$$
$$I_n(\theta) = -\frac{1}{n}\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\log f_\theta(X_i)$$

Using the Taylor expansion with mean value remainder we have

$$U_n(\theta_0) = U_n(T_n) + \frac{\partial}{\partial\theta}U_n(\theta)\Big|_{\theta=\tilde{T}_n}(\theta_0 - T_n) = U_n(T_n) - I_n(\tilde{T}_n)(\theta_0 - T_n)$$

for some $\tilde{T}_n$ between θ_0 and T_n. Rearranging this, we have

$$\begin{aligned} U_n(\theta_0) =& I_n(\tilde{T}_n)T_n + U_n(T_n) - I_n(\tilde{T}_n)\theta_0 \\ =& I_n(T_n)T_n + U_n(T_n) - I_n(T_n)\theta_0 \\ &+ [I_n(\tilde{T}_n) - I_n(T_n)](T_n - \theta_0). \end{aligned}$$

Moreover, we can then express this as

$$\sqrt{n}\{T_n + I_n(T_n)^{-1}U_n(T_n) - \theta_0\} = I_n(T_n)^{-1}\sqrt{n}U_n(\theta_0) - I_n(T_n)^{-1}[I_n(\tilde{T}_n) - I_n(T_n)]\sqrt{n}(T_n - \theta_0).$$

Noting that $I_n(T_n)^{-1} \to_p I_f(\theta_0)^{-1}$, $\sqrt{n}U_n(\theta_0) \to_d N(0, I_f(\theta_0))$, $I_n(\tilde{T}_n) - I_n(T_n) \to_p 0$, and $\sqrt{n}(T_n - \theta_0) \to N(0, \sigma^2(\theta_0))$ for some $\sigma^2(\theta_0)$, we have the RHS converges to $N(0, I_f(\theta_0)^{-1})$ by Slutsky's lemma. Hence, we have

$$\sqrt{n}(\hat{\theta}_n^{(1)} - \theta_0) = \sqrt{n}\{T_n + I_n(T_n)^{-1}U_n(T_n) - \theta_0\} \to_d N(0, I_f(\theta_0)^{-1}).$$

Lecture 8 (Confidence Intervals)

3. Dr. Jetson asked a random sample of 10000 UK households whether or not they own a robotic vacuum cleaner. She finds that 1300 of the households own a robotic vacuum and the other 8700 do not. Based on this data, she estimates that 13% of UK households own a robotic vacuum with a 95% confidence interval of 12.3% to 13.7%. Dr. Jetson tells you that

   > "There is a 95% probability that between 12.3% and 13.7% of UK households own a robotic vacuum cleaner."

   What is the main problem with the above statement? Provide a correct description of the confidence interval suitable for a non-statistician.

   Solution. The main problem with Dr. Jetson's statement is that it tries to make a probability statement based on the fixed interval and fixed proportion of UK households owning a robotic vacuum cleaner. The 0.95 probability refers to a property of the random confidence intervals. If Dr. Jetson repeated her survey many times and constructed 95% confidence intervals for the proportion each time, approximately 95% of the resulting intervals would contain the true proportion.

4. A random sample of 11 components in a factory is collected. The length in cm of each component is recorded below

   3.26 1.76 1.63 1.79 2.43 0.88 0.99 1.12 4.56 2.11 2.73

   Assume that the lengths are normally distributed with mean μ and variance σ^2. Construct a 99% confidence interval for μ.

   Solution. As pivotal quantity we use $\frac{\bar{x}-\mu}{s/\sqrt{n}}$ which is t_{n-1} distributed, where $n = 11$, where $\bar{x}$ and s^2 are the observed sample mean and variance. Using a table (or a calculator/computer), this implies

   $$P(|\frac{\bar{x} - \mu}{s^2/\sqrt{n}}| \leq k) = 0.99,$$

   where $k = 3.169$. Hence, a 99% confidence interval for μ is $(\bar{x} - s/\sqrt{n}k, \bar{x} + s/\sqrt{n}k) = (1.07, 3.16)$ (using $\bar{x} = 2.114545$ and $s^2 = 1.201427$).

5. Let $Y_1, \ldots, Y_n$ be i.i.d. Exp(λ), where $\lambda > 0$ is unknown.

   (a) Show that $2\lambda \sum_{i=1}^n Y_i$ has a χ^2-distribution with $2n$ degrees of freedom;

   (b) Derive a $(1 - \alpha) \times 100\%$ confidence interval for λ;

   (c) Using the following observations, compute a 95% confidence interval for λ.

1.04 1.39 0.1 2.04 4.73 0.89 0.51 0.89 0.66 0.93

(Note: for $X \sim \chi^2_{20}$, $P(X \leq 9.59) = 0.025$ and $P(X \leq 34.17) = 0.975$.)

Solution. Note that $2\lambda Y_i$ is Exp($\frac{1}{2}$), which is χ^2_2. Since the Y_i's are independent, $2\lambda \sum Y_i$ is the sum of n independent χ^2_2 random variables and is therefore distributed as χ^2_{2n}. Hence,

$$P(c_1 < 2\lambda \sum Y_i < c_2) = 1 - \alpha,$$

where $0 < \alpha < 1$ and where $P(X < c_1) = P(X > c_2) = \frac{\alpha}{2}$ and $X \sim \chi^2_{2n}$. A $1 - \alpha$ confidence interval for λ is thus given by $\left(\frac{c_1}{2\sum y_i}, \frac{c_2}{2\sum y_i}\right)$.

Hence, the confidence interval for the data based on $\sum Y_i$ is: $(0.36, 1.3)$

6. Find an approximate 95% confidence interval for the odds that a randomly selected UK household owns a robotic vacuum based on the data in Exercise 3. (Hint: use the delta method.)

Solution. A point estimate for the odds will be $0.13/0.87 = 0.1494253$.

Using the delta method, we know that with $Odds(p) = p/(1-p)$ and $Odds'(p) = 1/(1-p)^2$ the standard error for the odds will be

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n(1-\hat{p})^4}} = \sqrt{\frac{0.13(1-0.13)}{10000(1-0.13)^4}} = 0.004763577$$

which leads to an approximate 95% confidence interval with limits

$$0.1494253 \pm 1.96 \times 0.004763577$$

or $(0.1400887, 0.1587619)$.

Alternatively, since the odds are an invertible function of the probability, we can apply $p/(1-p)$ to the limits of the original interval and still have a valid 95% confidence interval. This leads to the very similar interval estimate

$$(0.1402509, 0.1587486)$$

7. Use the Bonferroni correction to find a 95% confidence region for (μ, σ^2) based on a random sample $X_1, \ldots, X_n$ from a $N(\mu, \sigma^2)$ distribution. Apply your result to construct a 95% confidence region for (μ, σ^2) based on the data in Exercise 4.

Solution. We will make use of example 27 from the lecture notes for a normal random sample with both μ and σ unknown. To obtain a 95% confidence region for (μ, σ^2) with the Bonferroni correction, we will construct two-sided (1-0.05/2)100% = 97.5% confidence intervals for each parameter.

Let $\bar{X}$ and S^2 be the sample mean and variance of the X_is.

The 97.5% confidence interval for μ has the form

$$\bar{X} \pm t_{n-1,0.9875}\frac{S}{\sqrt{n}} \implies \left(\bar{X} - t_{n-1,0.9875}\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1,0.9875}\frac{S}{\sqrt{n}}\right)$$

for $t_{n-1,0.9875}$ the value such that the $P(T_{n-1} \leq t_{n-1,0.9875}) = 0.9875 = 1 - 0.025/2$.

The 97.5% confidence interval for σ^2 has the form

$$\left(\frac{(n-1)S^2}{c_2}, \frac{(n-1)S^2}{c_1}\right)$$

where $P(\chi^2_{n-1} \leq c_1) = 0.0125$ and $P(\chi^2_{n-1} \leq c_2) = 0.9875$.

We have, by the Bonferroni correction, that

$$\left(\bar{X} - t_{n-1,0.9875}\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1,0.9875}\frac{S}{\sqrt{n}}\right) \times \left(\frac{(n-1)S^2}{c_2}, \frac{(n-1)S^2}{c_1}\right)$$

is a 95% confidence region for (μ, σ^2).

We use $\bar{x} = 2.114545$, $s^2 = 1.201427$, $n = 11$ to find that $t_{10,0.9875} = 2.633767$, $c_1 = 2.707213$, and $c_2 = 22.55825$. This results in the 95% confidence region (simultaneous confidence intervals)

$$(1.2441242.984967) \times (0.5325889, 4.4378752) \approx (1.2, 3.0) \times (0.5, 4.4).$$

R lab: The Bootstrap

This exercise introduces concepts through use of the R software package.

Let T_n be an asymptotically normal estimator of θ based on a random sample $Y_1, \ldots, Y_n$. We now consider a flexible method called the bootstrap that allows us to approximate the sampling distribution of T_n using observations $y_1, \ldots, y_n$.

The bootstrap sampling distribution can be used to construct confidence intervals for θ by either:

i. Computing $\mathrm{SE}(T_n)$ with respect to the bootstrap sampling distribution and using the formula $T_n \pm c_{\alpha/2}\,\mathrm{SE}(T_n)$ from the notes;

ii. Computing the $\alpha/2$ and $1-\alpha/2$ quantiles of the bootstrap sampling distribution.

This procedure is widely applicable, but is most useful for estimators where it is difficult to obtain a closed-form expression for $\mathrm{SE}(T_n)$.

In R, the code below shows how we usually compute $\bar{y}$ and estimate its standard error based on 4 data points: 2, 4, 9, and 12.

```
y <- c(2,4,9,12)
ybar <- mean(y)
se.ybar <- sqrt(var(y)/4)
```

Running the above code, we find that the standard error is about 2.29.

The bootstrap sampling distribution of $\bar{Y}$ is obtained by resampling the data points with replacement and computing $\bar{Y}$ based on the resampled data. There are 4 data points, so there are $4^4 = 256$ equally likely resamples. We can use R to obtain all 256 values in the bootstrap sampling distribution as follows:

```
# All 4^4 = 256 possible resamples with replacement
y.star <- expand.grid(y,y,y,y)

# All 256 sample means based on resampling w/replacement
ybar.star <- apply(y.star, 1, mean)

# The standard error based on this is
se.ybar.star <- sqrt(var(ybar.star))
```

From the above, we find that the bootstrap standard error is about 1.98.

It is a good idea to also visualise the bootstrap distribution of $\bar{Y}$. This can be achieve with `hist(ybar.star)`. For large sample sizes, we would expect the bootstrap sampling distribution to look approximately normal. The normal approximation for $n = 4$ seems to be less than ideal.

The number of bootstrap resamples n^n grows too quickly to be reasonable for the average statistician. Instead, we usually approximate the bootstrap sampling distribution by drawing a large number of random samples as follows.

```
set.seed(50011)

ybar.boot <- numeric(length = 10000)
for(i in 1:10000){
```

```
y.boot <- sample(y, size = 4, replace = TRUE)
ybar.boot[i] <- mean(y.boot)
}
se.ybar.boot <- sqrt(var(ybar.boot))
```

From the above, we find that the bootstrap standard error is about 1.99. This is nearly the same as the value obtained by enumerating all 256 samples.

8. Using the code examples above:

   (a) Construct three approximate 95% confidence intervals for the mean μ based on the formula $T_n \pm c_{\alpha/2}\,\mathrm{SE}(T_n)$ where the standard error is based on `se.ybar`, `se.ybar.star`, and `se.ybar.boot`.

   (b) Construct two additional approximate 95% confidence intervals for the mean with limits define by the 2.5% and 97.5% percentiles of `ybar.star` and `ybar.boot`. (Hint: use the `quantile()` function.)

   (c) Compare the similarities/differences in the confidence intervals you constructed in parts (a) and (b).

   (d) Replace the data `y` in your code with a random sample of $n = 30$ standard exponential random variables: `y <- rexp(n=30)`. Based on your previous code, construct an approximate 95% confidence interval for the mean based on a normal approximation. Construct two different bootstrap 95% confidence intervals for the mean based on 10000 resamples. (Note: you are not being asked to enumerate all 30^{30} resamples.)

Solution.

(a) We ran the following additional code in R to generate the 95% confidence intervals:

```
> ybar +c(-1,1)* 1.96*se.ybar
[1]  2.267995 11.232005
> ybar +c(-1,1)* 1.96*se.ybar.star
[1]  2.860867 10.639133
> ybar +c(-1,1)* 1.96*se.ybar.boot
[1]  2.848146 10.651854
```

So the three intervals are, to two decimals,

$$(2.27, 11.23), \qquad (2.86, 10.64), \qquad \text{and} \qquad (2.85, 10.65).$$

(b) We use the following code to generate the 95% confidence intervals based on the percentiles:

```
> quantile(ybar.star, c(0.025,0.975))
2.5% 97.5%
3.0  10.5
> quantile(ybar.boot, c(0.025,0.975))
2.5% 97.5%
3.0  10.5
```

So the two intervals are both $(3.0, 10.5)$ using this method.

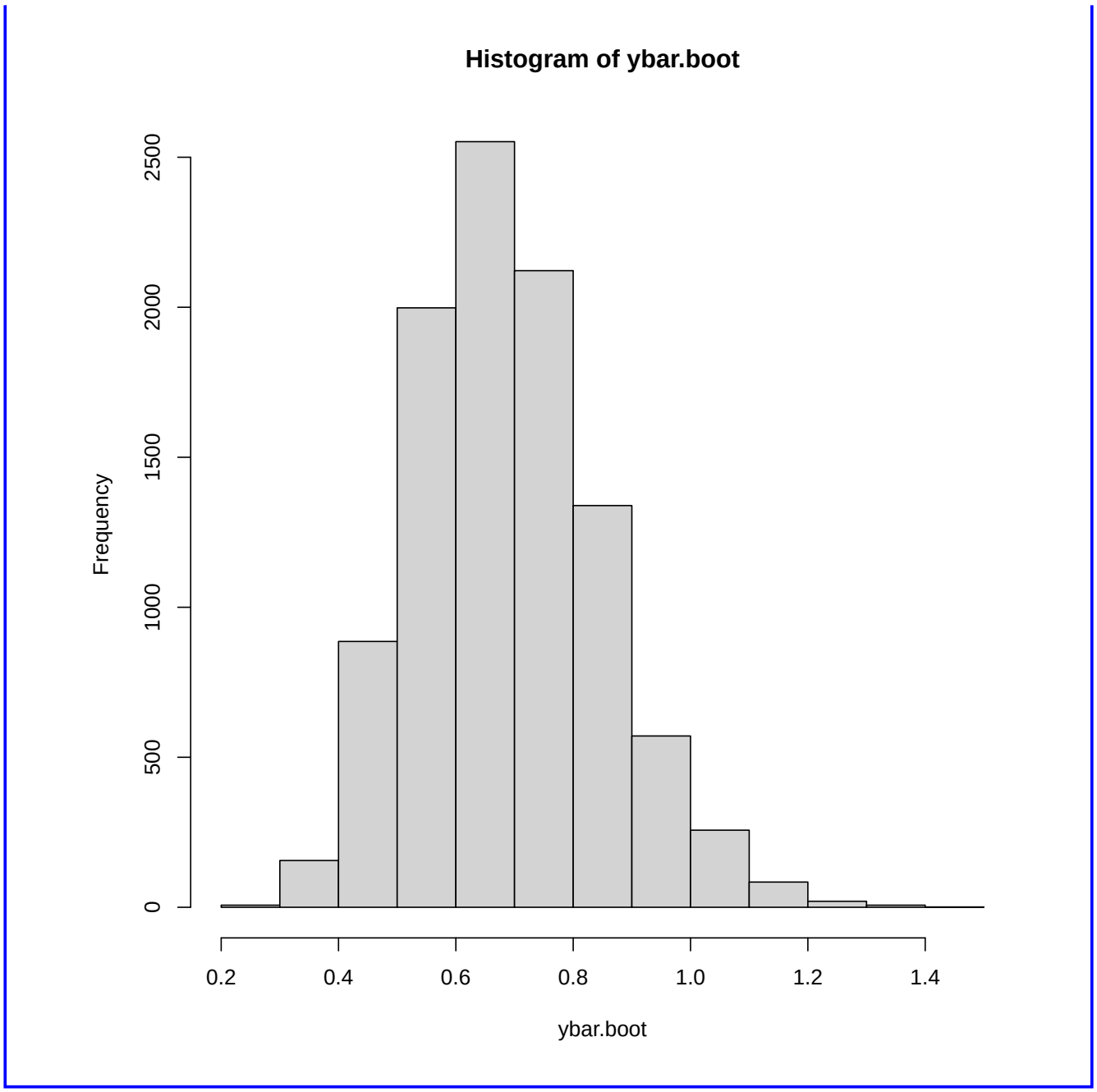
(c) The interval based on the usual standard error estimate $s/\sqrt{n}$ is wider than the bootstrap confidence intervals. The results of taking 10000 resamples do not differ greatly from enu-

merating the full sampling distribution. The percentile-based bootstrap confidence intervals result in the narrowest 95% confidence intervals in this example.

(d) When we replace the original data with the $n = 30$ random samples from the exponential distribution, we run the following code:

```
y <- rexp(n=30)
ybar <- mean(y)
se.ybar <- sqrt(var(y)/30)

# The bootstrap using Monte Carlo sampling
set.seed(50011)

ybar.boot <- numeric(length = 10000)
for(i in 1:10000){
y.boot <- sample(y, size = 30, replace = TRUE)
ybar.boot[i] <- mean(y.boot)
}
se.ybar.boot <- sqrt(var(ybar.boot))

# Confidence intervals using se.ybar, se.ybar.boot
# c_{\alpha/2} is approx 1.96
ybar +c(-1,1)* 1.96*se.ybar
ybar +c(-1,1)* 1.96*se.ybar.boot

# Confidence intervals using quantiles of ybar.boot
quantile(ybar.boot, c(0.025,0.975))
```

Note that your results may differ slightly depending on whether you reset your random seed. We obtain a 95% confidence interval of (0.3770655, 0.9975993) using the normal approximation, (0.3793849 0.9952799) using the bootstrap standard error, and (0.418570 1.031513) using the percentile method. Here, the first two intervals are fairly similar. The percentile-based bootstrap interval is shifted upward relative to the other two intervals. This may be because the percentiles are not exactly symmetric for our bootstrap distribution. The histogram of the resamples contained in `ybar.boot` supports this.


Histogram of ybar.boot
Frequency
0
500
1000
1500
2000
2500
0.2
0.4
0.6
0.8
1.0
1.2
1.4
ybar.boot