Imperial College London

Problem Sheet 3 Solutions

MATH50011
Statistical Modelling 1

Week 3

Lecture 5 (Asymptotic Normality)

1. Prove that if $X_1, X_2, \ldots$ converges in probability to X and h is a continuous function, then $h(X_1), h(X_2), \ldots$ converges in probability to $h(X)$.

Solution. Let $\epsilon > 0$ be given. We want to show that

$$\lim_{n\to\infty} P(|h(X_n) - h(X)| < \epsilon) = 1.$$

By continuity of h, we know that there exists $\delta \equiv \delta(\epsilon)$ such that

$$|X_n - X| < \delta \Rightarrow |h(X_n) - h(X)| < \epsilon.$$

This implies that

$$P(|X_n - X| < \delta) \leq P(|h(X_n) - h(X)| < \epsilon).$$

We know that the right hand side is bounded above by one. Taking limits we find

$$1 = \lim_{n\to\infty} P(|X_n - X| < \delta) \leq \lim_{n\to\infty} P(|h(X_n) - h(X)| < \epsilon) \leq 1,$$

where we have used $X_n \to_p X$. Hence, we conclude that $h(X_n) \to_p h(X)$.

2. Suppose that $X_1, \ldots, X_n$ are iid with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Define $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.

   (a) Show that S_n^2 is a consistent estimator of σ^2. Assume that all required higher order moments of X_i exist and are finite.

Solution. First, we note that

$$S_n^2 = \frac{n}{n-1}(U_n + V_n)$$

with $U_n = \frac{1}{n}\sum_{i=1}^n X_i^2$ and $V_n = -\bar{X}_n^2$. Since $U_n \to_p E(X^2)$ and $V_n \to_p -\mu^2$ by continuity (see problem 1), we have by Slutsky's lemma that $U_n + V_n \to_p E(X^2) - \mu^2 = \sigma^2$. Since $\frac{n}{n-1} \to_p 1$, further application of Slutsky's lemma leads to the desired conclusion $S_n^2 \to_p \sigma^2$.

(b) Use the result in (a) to show that

$$T_n = \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sqrt{S_n^2}}\right) \to_d N(0,1)\,.$$

Solution. By the CLT, we have that

$$Z_n = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \to_d N(0,1).$$

We can write

$$T_n = Z_n \frac{\sigma}{\sqrt{S_n^2}}.$$

By part (a) and continuity of $h(t) = \sigma/\sqrt{t}$ at $t \neq 0$, we have that $h(S_n^2) \to_p 1$. Hence, by Slutsky's lemma we have

$$T_n \to_d N(0,1).$$

3. Suppose that $X_1, \dots, X_n$ are iid strictly positive random variables with $E(\log X_i) = \mu$ and $\text{Var}(\log X_i) = \sigma^2$. Use the delta method to derive the asymptotic normality of the geometric mean $G_n = (\prod_{i=1}^n X_i)^{1/n}$.

Solution. Let $T_n = \log G_n = \frac{1}{n}\sum_{i=1}^n \log X_i$, which is the mean of iid random variables. By the CLT,

$$\sqrt{n}(T_n - \mu) \to_d N(0, \sigma^2).$$

We have that $G_n = \exp(T_n) = g(T_n)$ with $g(t) = g'(t) = exp(t)$. By the delta method,

$$\sqrt{n}(G_n - e^{\mu}) \to_d N(0, e^{2\mu}\sigma^2).$$

4. Suppose that $X_1, \dots, X_n$ are iid Uniform$(0, \theta)$ and define $T_n = \max(X_1, \dots, X_n)$. Find a sequence $a_n = n^k$ for some k such that $a_n(T_n - \theta) \to_d Z$. What is the distribution of Z?

Solution. From PS2 Q7, we know that $Var(T_n)$ is on the order of n^{-2}. To prevent $Var[a_n(T_n - \theta)] \to 0$, we might expect that $a_n = n^1 = n$ is an appropriate scaling factor.

In any case, $P(a_n(T_n - \theta) \le t) = P(T_n \le \theta + t/a_n) = \left(1 + \frac{t/\theta}{a_n}\right)^n$, where the probability is derived as in PS2 Q7. It is now evident that for $a_n = n$, as $n \to \infty$,

$$P(a_n(T_n - \theta) \le t) = \left(1 + \frac{t/\theta}{n}\right)^n \to e^{t/\theta}$$

for $t < 0$ and $P(a_n(T_n - \theta) \le t) = 1$ for all $t \ge 0$. That this sequence is supported on the negative reals follows immediately from noting that $P(T_n < \theta) = 1$ since $0 < X_i < \theta$ for each i.

5. Does $\sqrt{n}(T_n - \theta) \to_d N(0, \sigma^2)$ imply that T_n is consistent for θ? If yes, prove this. Otherwise, provide a counterexample.

Solution. We will make use of the following identity

$$\begin{aligned} P(|T_n - \theta| < \epsilon) &= P(-\epsilon < T_n - \theta < \epsilon) \\ &= P(T_n - \theta < \epsilon) - P(T_n - \theta \le -\epsilon) \\ &= P(\sqrt{n}(T_n - \theta) < \sqrt{n}\epsilon) - P(\sqrt{n}(T_n - \theta) \le -\sqrt{n}\epsilon). \end{aligned}$$

We want to show that for any $\delta > 0$, there exists n_0 such that for $n > n_0$ we have

$$P(|T_n - \theta| < \epsilon) > 1 - \delta.$$

Let $z > 0$ be such that $\Phi(z/\sigma) - \Phi(-z/\sigma) = 1 - \delta/2$, where $\Phi(t)$ denotes the standard normal cdf. Whenever $\sqrt{n}\epsilon > z \Leftrightarrow n > (z/\epsilon)^2$, we find (by the identity above) that

$$P(|T_n - \theta| < \epsilon) \ge P(\sqrt{n}(T_n - \theta) < z) - P(\sqrt{n}(T_n - \theta) \le -z).$$

By asymptotic normality, the right-hand side converges to $1-\delta/2$ for this choice of z. By definition of convergence, there exists a value n_1 such that for any $n > n_1$ the right-hand side is at least $1 - \delta$.

Taking $n_0 = \max(n_1, (z/\epsilon)^2)$, we have established that for $n > n_0$ we have

$$P(|T_n - \theta| < \epsilon) > 1 - \delta$$

as desired. This completes the proof.

6. Find the MLE for estimating θ based on a random sample $X_1, \ldots, X_n$ from the following distributions

(a) Bernoulli(θ); (see Example 8)

Solution. We have seen in the previous problem sheet that $\frac{\partial}{\partial\theta} \log f_\theta(x) = \frac{x}{\theta} - \frac{1-x}{1-\theta}$ for $n = 1$. Hence the MLE $\hat{\theta}_n$ solves

$$\sum_{i=1}^n \frac{\partial}{\partial\theta} \log f_\theta(X_i) = \sum_{i=1}^n \frac{X_i}{\theta} - \frac{1-X_i}{1-\theta} = \sum_{i=1}^n \frac{X_i - \theta}{\theta(1-\theta)} = 0.$$

Solving for θ we obtain the solution $\hat{\theta}_n = \bar{X}_n$, where $\bar{X}_n$ is the sample mean.
We have also previously shown that $\frac{\partial^2}{\partial\theta^2} \log f_\theta(x) = -\frac{x}{\theta^2} - \frac{1-x}{(1-\theta)^2} < 0$ which in turn implies $\hat{\theta}$ is indeed a point of maximum.

(b) Poisson(θ);

Solution. We have

$$f_\theta(x) = \frac{\theta^x e^{-\theta}}{x!}$$
$$\frac{\partial}{\partial\theta} \log f_\theta(x) = \frac{x}{\theta} - 1$$
$$\frac{\partial^2}{\partial\theta^2} \log f_\theta(x) = -\frac{x}{\theta^2}$$

so the MLE solves

$$\sum_{i=1}^n \frac{\partial}{\partial\theta} \log f_\theta(X_i) = \sum_{i=1}^n (\frac{X_i}{\theta} - 1) = 0$$

so that $\hat{\theta}_n = \bar{X}_n$. The second derivative with respect to θ is again negative, so that this is a point of maximum.

(c) Exponential(θ);

Solution. We have

$$\log f_\theta(x) = \log\theta - \theta x$$
$$\frac{\partial}{\partial\theta} \log f_\theta(x) = \frac{1}{\theta} - x$$
$$\frac{\partial^2}{\partial\theta^2} \log f_\theta(x) = -\frac{1}{\theta^2}$$

so the MLE solves

$$\sum_{i=1}^n \frac{\partial}{\partial\theta} \log f_\theta(X_i) = \sum_{i=1}^n (\frac{1}{\theta} - X_i) = 0$$

and $\hat{\theta}_n = 1/\bar{X}_n$. The second derivative with respect to θ is again negative, so that this is a point of maximum.

7. For the distributions in 6(a-c), find Z such that $\sqrt{n}(\hat{\theta}_n - \theta) \to_d Z$.

Solution. We know from the asymptotic normality of the MLE that in each case

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

where $I(\theta)$ is the Fisher information for sample of $n = 1$ individuals. From the previous problem sheet, we know

	Bernoulli	Poisson	Exponential
$I(\theta)$	$1/\theta(1-\theta)$	$1/\theta$	$1/\theta^2$
$I(\theta)^{-1}$	$\theta(1-\theta)$	θ	θ^2

We can also use the CLT directly to verify the distribution for the Bernoulli and Poisson distribution, since the MLE is also the sample mean. For the exponential distribution, the CLT can be applied in tandem with the delta method since the MLE is a differentiable function of the sample mean.

8. For the distributions in 6(a) and 6(b), find the MLE $\hat{\nu}_n$ of $\nu = g(\theta) = P_\theta(X_1 = 0)$ and show that $\sqrt{n}(\hat{\nu}_n - \nu) \to_d Z$. Find the distribution of Z in each case.

Solution. By invariance of the MLE, $\hat{\nu}_n = P_{\hat{\theta}_n}(X_1 = 0)$. For the Bernoulli distribution,

$$\hat{\nu} = 1 - \hat{\theta}$$

and $\sqrt{n}(\hat{\nu} - \nu) \to_d N(0, \theta(1-\theta))$ since $\hat{\nu} - \nu = -(\hat{\theta} - \theta)$.
For the Poisson distribution,

$$\hat{\nu} = e^{-\hat{\theta}}$$

and $\sqrt{n}(\hat{\nu}_n - \nu) = \sqrt{n}(e^{-\hat{\theta}_n} - e^{-\theta}) \to_d N(0, e^{-2\theta}\theta)$ by the delta method with $g(t) = e^{-t}$.

9. Suppose that we wish to estimate θ based on a random sample $X_1, \dots, X_n$ of Bernoulli(θ) random variables. However, we are only able to obtain a random sample $(Y_i, R_i), \dots, (Y_n, R_n)$ where the R_i's are iid Bernoulli(p_0) for known p_0 and $Y_i = R_i X_i$ for $i = 1, \dots, n$. Derive the MLEs $\hat{\theta}_a, \hat{\theta}_b$ and $\hat{\theta}_c$ for θ based on

   (a) The full data distribution of the X_i's;

   Solution. See, e.g., problem 6(a).

(b) The marginal distribution of the Y_i's;

Solution. Recall from the previous problem sheet that

$$\begin{aligned} f_\theta(y) &= [\theta p_0]^y (1-\theta p_0)^{1-y} \\ \frac{\partial}{\partial\theta} \log f_\theta(y) &= \frac{y}{\theta} - p_0 \frac{1-y}{1-\theta p_0} \\ \frac{\partial^2}{\partial\theta^2} \log f_\theta(y) &= -\frac{y}{\theta^2} - p_0^2 \frac{1-y}{(1-\theta p_0)^2} \\ I_n(\theta) &= -nE\left\{\frac{\partial^2}{\partial\theta^2} \log f_\theta(Y)\right\} = \frac{np_0}{\theta(1-\theta p_0)} \\ CRLB_Y &= \frac{\theta(1-\theta p_0)}{np_0} \end{aligned}$$

Hence, the MLE is the solution to

$$\sum_{i=1}^n \frac{\partial}{\partial\theta} \log f_\theta(Y_i) = \sum_{i=1}^n \frac{Y_i}{\theta} - p_0 \frac{1-Y_i}{1-\theta p_0} = 0$$

so that $\hat{\theta}_b = \frac{1}{np_0}\sum_{i=1}^n Y_i$. This is indeed a point of maximum since the second derivative is negative.

(c) The joint distribution of the (Y_i, R_i)'s.

Solution. Recall from the previous problem sheet that

$$\begin{aligned} f_\theta(y,r) &= p_0^r (1-p_0)^{1-r} \{\theta^y (1-\theta)^{1-y}\}^r \\ \frac{\partial}{\partial\theta} \log f_\theta(y,r) &= r\left[\frac{y}{\theta} - \frac{1-y}{1-\theta}\right] \\ \frac{\partial^2}{\partial\theta^2} \log f_\theta(y,r) &= -r\left[\frac{y}{\theta^2} + \frac{1-y}{(1-\theta)^2}\right] \\ I_n(\theta) &= -nE\left\{\frac{\partial^2}{\partial\theta^2} \log f_\theta(Y,R)\right\} = \frac{np_0}{\theta(1-\theta)} \\ CRLB_{Y,R} &= \frac{\theta(1-\theta)}{np_0} \end{aligned}$$

Hence, the MLE is the solution to

$$\sum_{i=1}^n \frac{\partial}{\partial\theta} \log f_\theta(Y_i, R_i) = \sum_{i=1}^n R_i \left[\frac{Y_i}{\theta} - \frac{1-Y_i}{1-\theta}\right] = 0$$

so that $\hat{\theta}_c = \sum_{i=1}^n Y_i / \sum_{i=1}^n R_i$. This is indeed a point of maximum since the second derivative is negative.

10. Let T_n and U_n be estimators of θ such that

$$\begin{aligned} \sqrt{n}(T_n - \theta) &\to_d N(0, \sigma_T^2) \\ \sqrt{n}(U_n - \theta) &\to_d N(0, \sigma_U^2)\,. \end{aligned}$$

The *asymptotic relative efficiency* of T_n with respect to U_n is σ_T^2/σ_U^2.

Find the asymptotic distributions of the MLEs in 8(b) and 8(c) and calculate the asymptotic relative efficiency of $\hat{\theta}_b$ to $\hat{\theta}_c$. Which of the MLEs do you prefer for estimating θ? Quantify the loss in efficiency of your preferred estimator to $\hat{\theta}_a$ that is based on the (unobserved) X_i's. Explain.

Solution. Noting that the asymptotic variance of the MLE is the CRLB for $n = 1$ we use calculations of the previous problem sheet below.

In particular, we have that

$$\sqrt{n}(\hat{\theta}_b - \theta) \to_d N(0, \theta(1-\theta p_0)/p_0)$$

and

$$\sqrt{n}(\hat{\theta}_c - \theta) \to_d N(0, \theta(1-\theta)/p_0).$$

The asymptotic relative efficiency of $\hat{\theta}_b$ to $\hat{\theta}_c$ is

$$\frac{\theta(1-\theta p_0)/p_0}{\theta(1-\theta)/p_0} = \frac{1-\theta p_0}{1-\theta} \geq 1$$

with equality iff $p_0 = 1$ (so that X_i is observed with probability 1). Hence, we prefer the MLE $\hat{\theta}_c$ based on the joint distribution of (Y, R) on the basis of the asymptotic relative efficiency.

The asymptotic relative efficiency of $\hat{\theta}_c$ to the "complete data" MLE $\hat{\theta}_a$ is

$$\frac{\theta(1-\theta)/p_0}{\theta(1-\theta)} = \frac{1}{p_0} \geq 1$$

with equality iff $p_0 = 1$ (so that $R_i = 1$ and X_i is observed with probability 1).

Roughly speaking, a sample of (Y_i, R_i)s provides only a fraction of the information about θ that direct observation of the X_is would. This fraction is precisely equal to p_0.

$\texttt{R}$ lab: One-Step Estimators

This exercise is intended to reinforce concepts through use of the $\texttt{R}$ *software package.*

In the notes, we saw that numerical methods can facilitate maximisation of the (log) likelihood. In this lab, we illustrate how a simple one-step update to an initial estimator can lead to an accurate approximation of the MLE. The step we take is based on Newton's method.

Suppose that $X_1, \ldots, X_n$ are iid with pdf $f_\theta(x)$. Define

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

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

The one-step estimator is defined as $\hat{\theta}_n^{(1)} = T_n - I_n(T_n)^{-1}U_n(T_n)$, where T_n is an initial estimator of θ. If T_n is an asymptotically normal estimator of θ, then

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

You will prove this in the next problem sheet.

11. In this exercise, you will implement a simulation study to explore the behavior of the one-step estimator for the location parameter θ of the Cauchy(θ) distribution with pdf

$$f_\theta(x) = \frac{1}{\pi\left[1 + (x-\theta)^2\right]} \quad -\infty < x < \infty, -\infty < \theta < \infty.$$

Note that $f_\theta(x)$ is symmetric about θ. However, $E_\theta(X)$ does not exist for the Cauchy distribution so the sample mean would be an awful estimator here. Instead, we will use the sample median as an initial estimator of θ.

After drawing $X_1, \ldots, X_n$ i.i.d. Cauchy(θ), the sample median $\hat{m}_n$ will be computed and stored as an initial estimator. The values of $U_n(\hat{m}_n)$ and $I_n(\hat{m}_n)$ are then computed and used to construct a one-step estimator $\hat{\theta}_n^{(1)}$ based on $\hat{m}_n$. This experiment will be independently replicated a total of 1000 times, so that we can approximate the sampling distributions of $\hat{m}_n$ and $\hat{\theta}_n^{(1)}$.

The R code below implements the simulation study for $n = 10$ and $\theta = 0$.

```
set.seed(50011)
result.m <- logical(length = 1000)
result.t1 <- logical(length = 1000)
n <- 10
theta <- 0
for(i in 1:1000){
   X <- rcauchy(n, location = 0)
   m <- median(X)
   U <- NULL
   I <- NULL
   t1 <- m - U/I
   result.m[i] <- sqrt(n)*(m-theta)
   result.t1[i] <- sqrt(n)*(t1-theta)
}
hist(result.m, freq=FALSE)
hist(result.t1, freq=FALSE)
```

Note that the command `set.seed(50011)` ensures that you obtain the same results each time you run this set of commands.

Type the above commands into an R script and then:

(a) Derive expressions for $U_n(\hat{m}_n)$ and $I_n(\hat{m}_n)$ in terms of `X` and `m`. Use your expressions to replace the appropriate `NULL` definitions in the for loop.

Solution. Taking derivatives of the log-likelihood we find

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

and

$$I_n(\theta) = -\frac{\partial}{\partial\theta} U_n(\theta) = -\sum_{i=1}^{n} 2\frac{1 - (X_i - \theta)^2}{[1 + (X_i - \theta)^2]^2}.$$

In the code above, we can assign

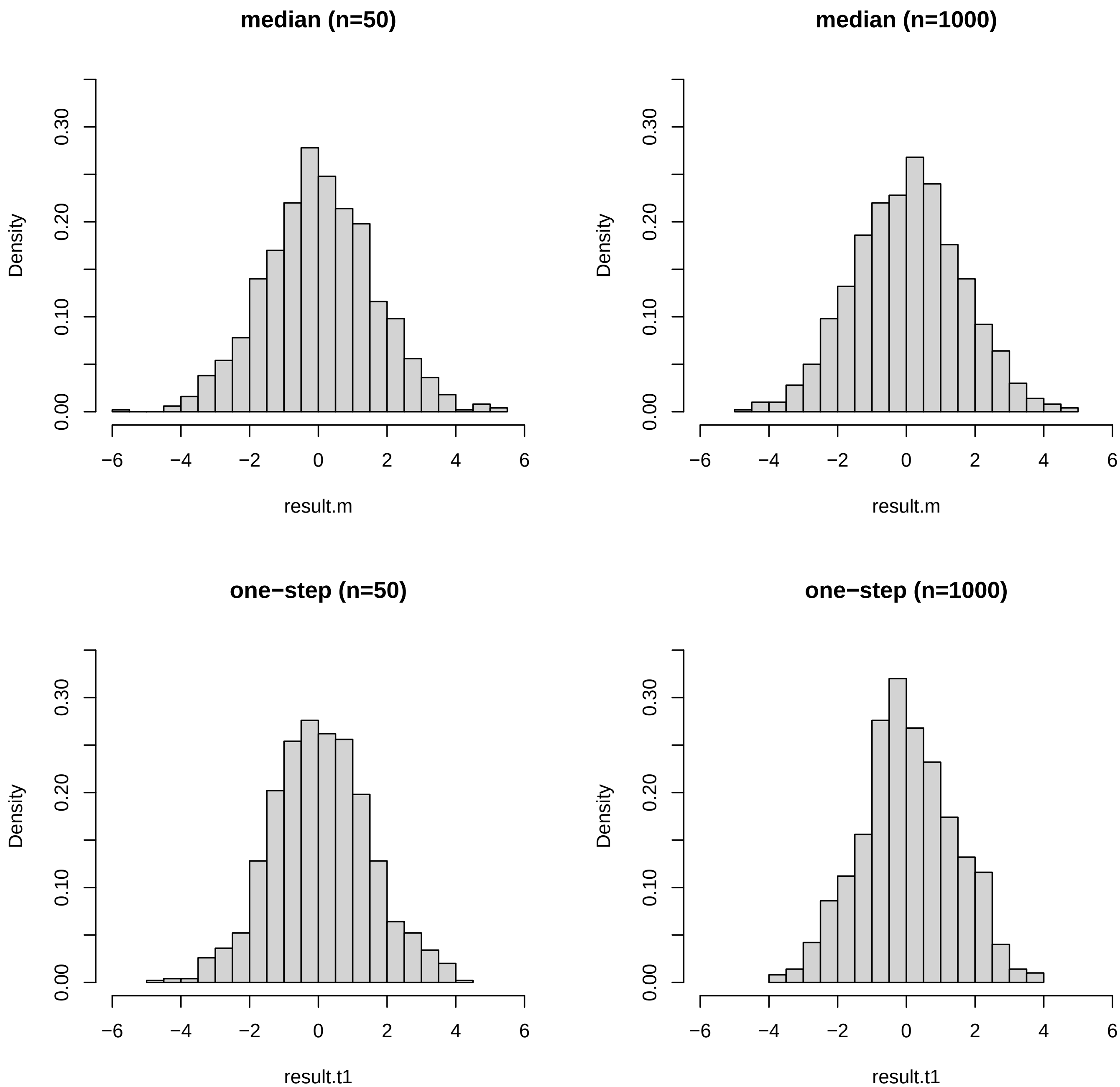
```
U <- 2*sum((X-m)/(1+(X-m)^2))
I <- -2*sum((1-(X-m)^2)/(1+(X-m)^2)^2)
```

(b) Comment on why it is reasonable to store the values of $\sqrt{n}(\hat{m}_n - \theta)$ and $\sqrt{n}(\hat{\theta}_n^{(1)} - \theta)$ instead of $\hat{\theta}_n^{(1)}$ and $\hat{m}_n$.

Solution. We are concerned about convergence of the scaled and centered sequences. We can always solve for the estimators based on the stored values.

(c) Explore how each histogram changes by increasing the value of `n` in this code to, e.g. $n = 30, 50, 100, 200, 500, 1000$. You might also compare other, say numerical, summaries (e.g. mean, variance, quantiles).

Solution. The histograms you generate should suggest that the median has a sampling distribution with slightly higher spread. See below for examples for $n = 50$ and $n = 1000$. In particular, for $n = 1000$, all of the bins for the one-step estimator are contained in the interval [-4,4] whereas the histogram for the median extends beyond this interval.



(d) Referring to your results from (c), comment on whether you prefer the sample median or one-step estimator for estimating θ in this setting.

Solution. Using the histograms from the 1000 simulation experiments as approximations to the sampling distribution, both histograms appear to be centered near zero for large sample sizes. However, the one-step estimator is less variable in larger sample sizes. We would prefer the one-step estimator based on these observations.

Challenge Do your simulations provide evidence that $\sqrt{n}(\hat{\theta}_n^{(1)} - \theta)$ convergences in distribution to a $N(0, I_f(\theta)^{-1})$ random variable? Explain your answer using appropriate graphical and/or numerical evidence.

Solution. The histogram for $n = 1000$ may not look immediately like a normal distribution (it is not quite symmetric or bell-shaped), but this may be due to Monte Carlo error. We can also overlay the density of a normal distribution to help with our assessment.

See Figure 1 below for 9 independent replications of the experiment. Repeating the experiment more than 1000 times (changing the definition of the loop and results vector) could be necessary to get a reliable picture of the sampling distribution. In Figure 2, we see that by increasing the number of replications of the experiment to 9000 there is much less variability between each histogram (and each histogram fits fairly well to the $N(0, I(\theta)^{-1})$ density).

Simulation studies such as this one are common in statistical research. The Stochastic Simulations module in Year 3 explores such ideas in greater detail.

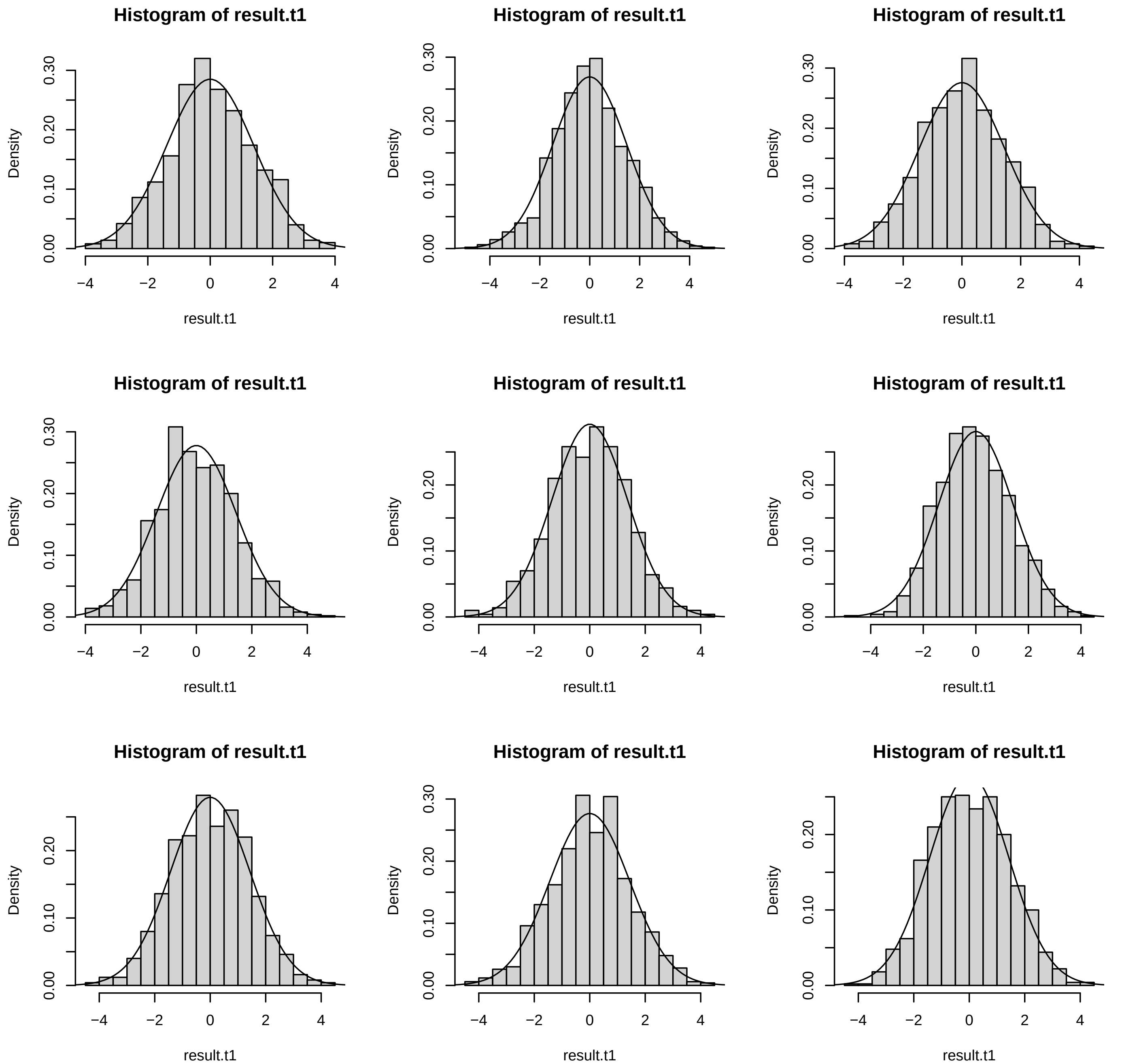


Figure 1: Nine histograms generated from independent runs with 1000 replications of the $n = 1000$ Cauchy one-step estimator.

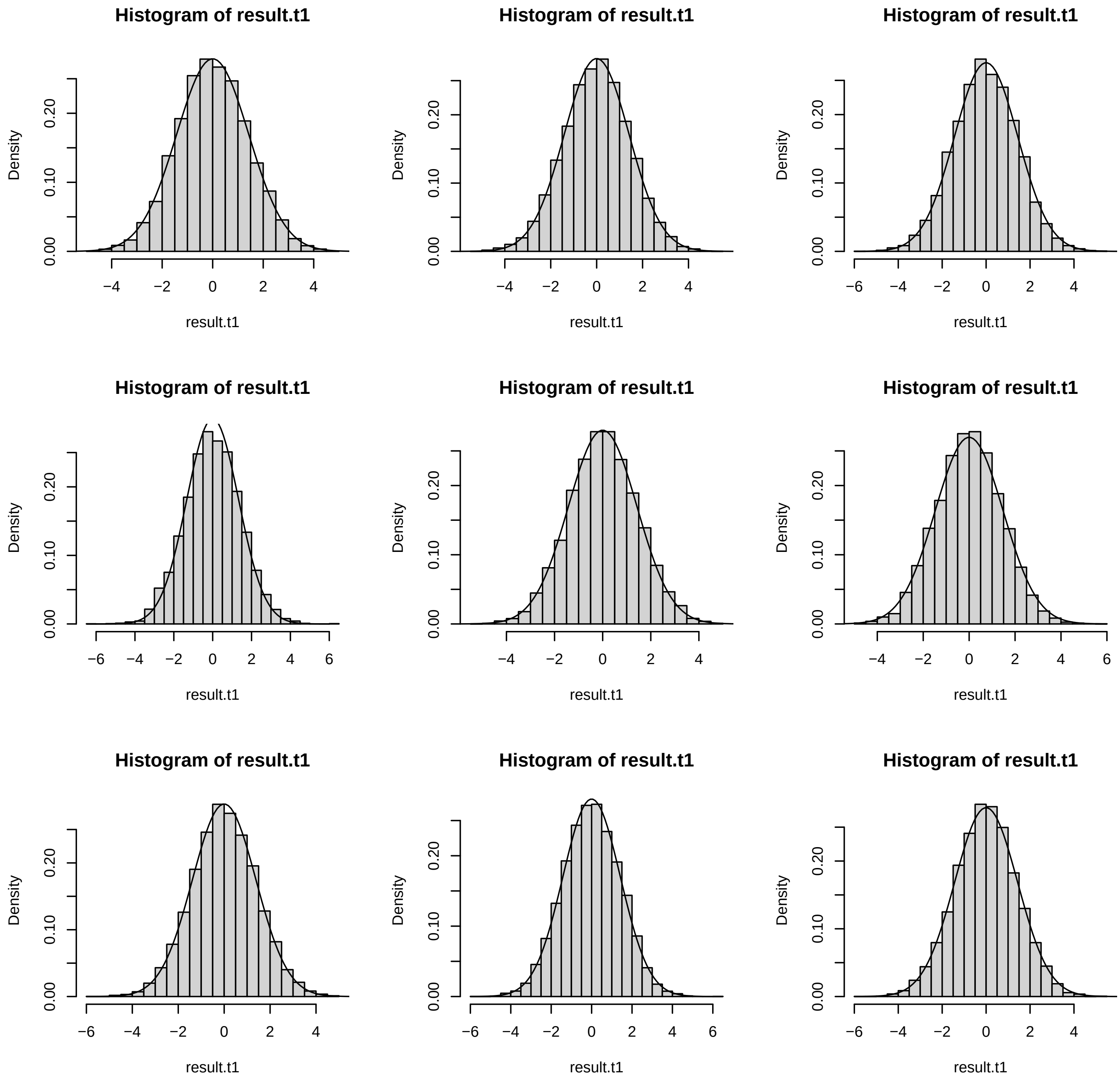


Figure 2: Nine histograms generated from independent runs with 9000 replications of the $n = 1000$ Cauchy one-step estimator.