1

Introduction

...where a motivation to study uniform laws of large numbers and central limit theorems is given. The examples concern parametric and nonparametric models, and contain indications on how we plan to derive consistency, rates of convergence and asymptotic normality of estimators.

Let $X_1, X_2, \ldots$ be independent copies of a random variable X with distribution P on $(\mathscr{X}, \mathscr{A})$. If $(\mathscr{X}, \mathscr{A})$ is the real line, equipped with a Borel σ-algebra, then by the strong law of large numbers, the sample mean $\bar{X} = (1/n)\sum_{i=1}^n X_i$ of the first n observations converges almost surely to the population mean $\mu = EX$, as $n \to \infty$. If moreover X has finite variance σ^2, the central limit theorem states that for n large, $\bar{X}$ is approximately normally distributed with mean μ and variance σ^2/n. For a general measurable space $(\mathscr{X}, \mathscr{A})$, such results hold for the sample mean $(1/n)\sum_{i=1}^n g(X_i)$, where $g : \mathscr{X} \to \mathbf{R}$ is some (measurable) real-valued transformation. This observation will be our starting point.

Strong law of large numbers *If $Eg(X)$ exists, then*

(1.1) $$\frac{1}{n}\sum_{i=1}^n g(X_i) \to Eg(X), \text{ a.s.}.$$

Central limit theorem *If $\sigma_g^2 = \mathrm{var}(g(X))$ exists, then*

(1.2) $$\frac{1}{\sqrt{n}}\sum_{i=1}^n (g(X_i) - Eg(X)) \to^{\mathscr{L}} \mathscr{N}(0, \sigma_g^2).$$

For example, if we take g as the indicator function 1_A of some set $A \in \mathscr{A}$, (1.1) states the convergence of the proportion of observations falling into the set A, to the probability of the set A. Result (1.2) is then the classical central limit theorem for Bernoulli random variables.

The convergence in (1.1) and (1.2) holds for a *fixed* function g, but can be extended to hold for several g simultaneously. Consider a class $\mathscr{G} = \{g_\theta : \theta \in \Theta\}$ of functions on $\mathscr{X}$, indexed by a parameter θ in a metric space Θ. Assume that $Eg_\theta(X)$ and $\text{var}(g_\theta(X))$ exist for all $\theta \in \Theta$. We shall investigate to what extent (1.1) and (1.2) hold *uniformly* in $\theta \in \Theta$. Although the study of uniform convergence is of interest in itself, our motivation is given by the numerous applications in asymptotic statistics. Let $\{\hat{\theta}_n\}$ be some (possibly) random sequence in Θ, converging to $\theta_0 \in \Theta$ (a.s. or in probability). We have in mind the situation where $\hat{\theta}_n$ is an estimator of θ_0, based on the first n observations. Uniform results would lead to something like a law of large numbers and a central limit theorem for $g_{\hat{\theta}_n}$. In the next section, we present three examples to illustrate the importance of this. The first example sets out with a simple parametric model, where one sees that extensions of (1.1) and (1.2) can be used to prove asymptotic normality of the maximum likelihood estimator. Next, we note that a parametric model is not always appropriate. To indicate how to proceed in nonparametric models, we place ourselves in a general context in Example 1.2. We indicate there how extensions of (1.1) play a role in the development of a general theory on consistency of maximum likelihood estimators. The last example in this chapter shows that also in nonparametric models, one can exploit extensions of (1.2) to arrive at asymptotic normality.

1.1. Some examples from statistics

Example 1.1. A binary choice model Let $X_i = (Y_i, Z_i)$, with Z_i the education of individual i, $Y_i = 1$ if individual i has a job, and $Y_i = 0$ otherwise. Suppose that we asked n individuals for their education and employment status. We want to model how the probability of having a job depends on education, and estimate the parameters in the model.

Case (i) The logit model is

$$P(Y = 1 \mid Z = z) = F_0(\alpha_0 + \theta_0 z),$$

with $F_0(\xi) = e^\xi/(1+e^\xi)$ the distribution function of the logistic distribution. To avoid notational digressions, we assume that only the parameter $\theta_0 \in \mathbf{R}$ is unknown, and that $\alpha_0 = 0$.

Denote the maximum likelihood estimator of θ_0 by $\hat{\theta}_n$, i.e. $\hat{\theta}_n$ is the value of θ that maximizes the (conditional) log-likelihood

$$\sum_{i=1}^{n} \log p_\theta(Y_i \mid Z_i),$$

where

$$p_\theta(y \mid z) = F_0^y(\theta z)(1 - F_0(\theta z))^{1-y}.$$

To investigate the asymptotic behaviour of this estimator, we introduce the score function

$$l_\theta(y, z) = \frac{d}{d\theta} \log p_\theta(y \mid z) = z(y - F_0(\theta z)),$$

and let

$$g_\theta(z) = \begin{cases} -\dfrac{l_\theta(y,z) - l_{\theta_0}(y,z)}{\theta - \theta_0} = z\dfrac{F_0(\theta z) - F_0(\theta_0 z)}{\theta - \theta_0}, & \text{if } \theta \neq \theta_0, \\ z^2 F_0(\theta_0 z)(1 - F_0(\theta_0 z)), & \text{if } \theta = \theta_0 \,. \end{cases}$$

Note that as $\theta \to \theta_0$, $g_\theta(z) \to g_{\theta_0}(z)$ for all z. Lemma 1.1 below assumes that we have some type of law of large numbers for $g_{\hat{\theta}_n}$.

Lemma 1.1 *Suppose that*

(1.3) $$\frac{1}{n}\sum_{i=1}^{n} g_{\hat{\theta}_n}(Z_i) \to_{\mathbf{P}} E g_{\theta_0}(Z) := I_{\theta_0}.$$

where $I_{\theta_0} > 0$.

Then $\sqrt{n}(\hat{\theta}_n - \theta_0)$ *is asymptotically* $\mathcal{N}(0, 1/I_{\theta_0})$*-distributed.*

Proof Since $\hat{\theta}_n$ maximizes the likelihood, the derivative of the log-likelihood at $\hat{\theta}_n$ is zero:

(1.4) $$\sum_{i=1}^{n} l_{\hat{\theta}_n}(Y_i, Z_i) = 0.$$

Clearly,

(1.5) $$\sum_{i=1}^{n} l_{\hat{\theta}_n}(Y_i, Z_i) = \sum_{i=1}^{n} l_{\theta_0}(Y_i, Z_i) - (\hat{\theta}_n - \theta_0)\sum_{i=1}^{n} g_{\hat{\theta}_n}(Z_i).$$

Combine (1.4) and (1.5) to give

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_{\theta_0}(X_i)}{\frac{1}{n}\sum_{i=1}^{n} g_{\hat{\theta}_n}(Z_i)};$$

recall that $X_i = (Y_i, Z_i)$. Note that

$$El_{\theta_0}(X) = 0, \quad \text{var}\big(l_{\theta_0}(X)\big) = I_{\theta_0}. \tag{1.6}$$

By the central limit theorem, the numerator is asymptotically $\mathcal{N}(0, I_{\theta_0})$-distributed, so the result follows immediately from assumption (1.3). □

Assumption (1.3) corresponds to an extension of the law of large numbers. Another approach towards proving asymptotic normality is based on an extension of the central limit theorem. To formulate this idea, we introduce the quantities

$$m_\theta = El_\theta(X),$$

and

$$\sigma_\theta^2 = \text{var}(l_\theta(X)).$$

For each θ, $\sum_{i=1}^n (l_\theta(X_i) - m_\theta)/\sqrt{n}$ converges weakly to a $\mathcal{N}(0, \sigma_\theta^2)$-distribution. If we consider $\sum_{i=1}^n (l_\theta(X_i) - m_\theta)/\sqrt{n}$ as a stochastic process indexed by θ, one can think of condition (1.7) below as asymptotic continuity of this process at θ_0.

We shall make use of the stochastic order symbol $o_{\mathbf{P}}(1)$. For example $\hat{\theta}_n = \theta_0 + o_{\mathbf{P}}(1)$ means that $\hat{\theta}_n$ converges to θ_0 in probability. For a formal definition, see Section 2.1. The general formulation of Lemma 1.2, for M-estimators of a finite-dimensional parameter, can be found in Section 12.3, Lemma 12.7.

Lemma 1.2 *Suppose that $\hat{\theta}_n = \theta_0 + o_{\mathbf{P}}(1)$, and that*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(l_{\hat{\theta}_n}(X_i) - m_{\hat{\theta}_n} \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (l_{\theta_0}(X_i) - m_{\theta_0}) + o_{\mathbf{P}}(1). \tag{1.7}$$

Then $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically $\mathcal{N}(0, 1/I_{\theta_0})$-distributed.

Proof Note first that

$$m_{\theta_0} = 0, \ \sigma_{\theta_0}^2 = I_{\theta_0},$$

and moreover that

$$\left. \frac{d}{d\theta} m_\theta \right|_{\theta=\theta_0} = -I_{\theta_0}. \tag{1.8}$$

Again, we rewrite the derivative of the log-likelihood at $\hat{\theta}_n$:

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^n l_{\hat{\theta}_n}(X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(l_{\hat{\theta}_n}(X_i) - m_{\hat{\theta}_n} \right) + \sqrt{n} m_{\hat{\theta}_n}.$$

Using assumption (1.7), we see that

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(l_{\theta_0}(X_i) - m_{\theta_0} \right) + o_{\mathbf{P}}(1) + \sqrt{n} m_{\hat{\theta}_n}$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l_{\theta_0}(X_i) + o_{\mathbf{P}}(1) - \sqrt{n}(\hat{\theta}_n - \theta_0)(I_{\theta_0} + o_{\mathbf{P}}(1)),$$

where in the last step, we invoked $m_{\theta_0} = 0$ and (1.8). The result now follows from the central limit theorem for l_{θ_0}. □

Case (ii) In Case (i), we assumed a parametric model for the probability of a job, given education, but there appears to be no intrinsic reason why the probability of having a job depends on education in this specific way. All we know is for instance that the higher the education, the more likely it is to have a job. In that case the model would be

$$P(Y = 1 \mid Z = z) = F_0(z),$$

with F_0 any increasing function of z satisfying $0 \le F_0(z) \le 1$. The parameter space is now

$$\Lambda = \{F : \mathbf{R} \to [0, 1],\ F \text{ increasing}\}.$$

The maximum likelihood estimator $\hat{F}_n$ is that value of $F \in \Lambda$ that maximizes

$$\sum_{i=1}^{n} \left(Y_i \log F(Z_i) + (1 - Y_i) \log\left(1 - F(Z_i)\right) \right). \tag{1.9}$$

It is no longer so easy to apply the argument that the derivative of the log-likelihood at $\hat{F}_n$ is zero. After all, what is a derivative in this situation? (See the first example in Section 11.2.3 for more details.) Moreover, how do we measure the distance between $\hat{F}_n$ and F_0? There are several possibilities here. In the next example, we shall indicate how one can prove consistency in the so-called Hellinger metric by employing a uniform law of large numbers. One can also think of using the $L_2(Q)$-distance, with Q the distribution of Z:

$$\|\hat{F}_n - F_0\|_Q = \left(\int \left(\hat{F}_n(z) - F_0(z) \right)^2 dQ(z) \right)^{1/2}$$

It will be shown in Example 7.4.3 that $\|\hat{F}_n - F_0\|_Q = O_{\mathbf{P}}(n^{-1/3})$ (for an explanation of stochastic order symbols, see Section 2.1). The same rate holds true for the Hellinger metric. Compare this with Case (i), where the rate of convergence for $\hat{\theta}_n$ is $O_{\mathbf{P}}(n^{-1/2})$. So the price one has to pay for not assuming the parametric model is that the rate of convergence is much slower. We shall present a quantification of this phenomenon. The 'size' or 'richness' of parameter space will be measured by means of its *entropy* (see Section 2.3 for a definition), and this entropy will be used to calculate the rate of convergence (see for example Theorem 7.4).

Case (iii) One can think of many intermediate models between Case (i) and Case (ii). Here is an example. Suppose that the probability of having a job indeed increases with education, but the amount of increase is lower at higher education levels. This means that $P(Y = 1 \mid Z = z)$ is a concave function of z:

$$P(Y = 1 \mid Z = z) = F_0(z),$$

with

$$F_0 \in \tilde{\Lambda} = \left\{ F : \mathbf{R} \to [0,1],\ 0 \le \frac{dF(z)}{dz} \le M,\ F(z) \text{ concave} \right\}.$$

The log-likelihood is of the same form as in Case (ii), but we maximize it over a smaller parameter space $\tilde{\Lambda}$ in order to get the maximum likelihood estimator $\hat{F}_n$. It turns out that under regularity conditions on Q, the rate of convergence is $O_{\mathbf{P}}(n^{-2/5})$ (see Example 7.4.3). This is an improvement as compared to Case (ii), due to the fact that we assumed a parameter space with smaller *entropy*.

Example 1.2. Maximum likelihood In this example, we study the maximum likelihood problem in general terms. Let X have density $p_{\theta_0}(x)$, $\theta_0 \in \Theta$, with respect to a σ-finite measure μ. Here Θ may be finite-dimensional (as was the case in Example 1.1, Case (i)) or infinite-dimensional (as in Example 1.1, Case (ii) and (iii)). The maximum likelihood estimator $\hat{\theta}_n$ maximizes the log-likelihood $\sum_{i=1}^n \log p_\theta(X_i)$ over all $\theta \in \Theta$. The idea here is that the log-likelihood will be close to its expectation for large sample sizes. Because the expected log-likelihood is maximized by θ_0, $\hat{\theta}_n$ is indeed a sensible estimator. What we need to turn this idea into a rigorous argument is an extension of the law of large numbers (1.1).

The estimator $\hat{\theta}_n$ maximizes the likelihood over $\theta \in \Theta$. Because $\theta_0 \in \Theta$, we therefore have

$$\sum_{i=1}^n \log \frac{p_{\theta_0}(X_i)}{p_{\hat{\theta}_n}(X_i)} \le 0. \tag{1.10}$$

On the other hand, for all θ,

$$E \log \frac{p_{\theta_0}(X)}{p_\theta(X)} \ge 0, \tag{1.11}$$

(see Problem 1.3), with equality if $\theta = \theta_0$. The quantity given in (1.11) is the Kullback–Leibler information

$$K(p_\theta, p_{\theta_0}) = \int \left(\log \frac{p_{\theta_0}}{p_\theta}\right) p_{\theta_0}\, d\mu.$$

Let

$$g_\theta = \log \frac{p_{\theta_0}}{p_\theta},$$

so that

$$K(p_\theta, p_{\theta_0}) = E g_\theta(X).$$

Then by (1.10),

$$0 \geq \frac{1}{n}\sum_{i=1}^n g_{\hat\theta_n}(X_i) = \frac{1}{n}\sum_{i=1}^n g_{\hat\theta_n}(X_i) - K(p_{\hat\theta_n}, p_{\theta_0}) + K(p_{\hat\theta_n}, p_{\theta_0}),$$

or,

$$K(p_{\hat\theta_n}, p_{\theta_0}) \leq \left|\frac{1}{n}\sum_{i=1}^n g_{\hat\theta_n}(X_i) - K(p_{\hat\theta_n}, p_{\theta_0})\right|. \tag{1.12}$$

By the law of large numbers, for each θ,

$$\left|\frac{1}{n}\sum_{i=1}^n g_\theta(X_i) - K(p_\theta, p_{\theta_0})\right| \to 0, \quad \text{a.s.}$$

Suppose this is also true for the sequence $\{\hat\theta_n\}$:

$$\left|\frac{1}{n}\sum_{i=1}^n g_{\hat\theta_n}(X_i) - K(p_{\hat\theta_n}, p_{\theta_0})\right| \to 0, \quad \text{a.s.} \tag{1.13}$$

Then it follows from (1.12) that the Kullback–Leibler information converges to zero. The Kullback–Leibler information is not a distance function, but its convergence often implies consistency of $\hat\theta_n$ in some metric of interest. A convenient metric is the Hellinger metric, defined as

$$h(p_\theta, p_{\theta_0}) = \left(\frac{1}{2}\int (p_\theta^{1/2} - p_{\theta_0}^{1/2})^2\, d\mu\right)^{1/2}.$$

The next lemma shows that convergence of the Kullback–Leibler information always yields consistency in the Hellinger metric.

Lemma 1.3 *We have*

$$h^2(p_\theta, p_{\theta_0}) \leq \frac{1}{2} K(p_\theta, p_{\theta_0}).$$

Proof Use the fact that $(1/2)\log v \leq v^{1/2} - 1$ for all $v > 0$:

$$\frac{1}{2} \log \frac{p_\theta(x)}{p_{\theta_0}(x)} \leq \frac{p_\theta^{1/2}(x)}{p_{\theta_0}^{1/2}(x)} - 1,$$

so

$$\frac{1}{2} K(p_\theta, p_{\theta_0}) \geq 1 - E\left(\frac{p_\theta^{1/2}(X)}{p_{\theta_0}^{1/2}(X)}\right).$$

Observe that

$$1 - E\left(\frac{p_\theta^{1/2}(X)}{p_{\theta_0}^{1/2}(X)}\right) = 1 - \int p_\theta^{1/2} p_{\theta_0}^{1/2}\, d\mu,$$

and since a density integrates to one,

$$\begin{aligned} 1 - \int p_\theta^{1/2} p_{\theta_0}^{1/2}\, d\mu &= \frac{1}{2} \int p_\theta\, d\mu + \frac{1}{2} \int p_{\theta_0}\, d\mu - \int p_\theta^{1/2} p_{\theta_0}^{1/2}\, d\mu \\ &= \frac{1}{2} \int (p_\theta^{1/2} - p_{\theta_0}^{1/2})^2\, d\mu = h^2(p_\theta, p_{\theta_0}). \end{aligned}$$

□

In summary, the maximum likelihood estimator $\hat{\theta}_n$ maximizes an empirical average, whereas θ_0 maximizes the expectation. If averages converge to expectations in a broad enough sense, this implies consistency of $\hat{\theta}_n$. We remark here that (1.13) is sometimes difficult to prove and perhaps not true. However, a modification of the idea presented here can demonstrate consistency in Hellinger distance, although one then might lose the convergence of the Kullback–Leibler information (see Section 4.1).

Also the rate at which the maximum likelihood estimator converges can be obtained along these lines. For each θ

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (g_\theta(X_i) - K(p_\theta, p_{\theta_0})) \to^{\mathscr{L}} \mathscr{N}(0, \sigma_\theta^2), \tag{1.14}$$

provided $\sigma_\theta^2 = \mathrm{var}(g_\theta(X)) < \infty$. First of all, this implies that for each such θ,

$$\left| \frac{1}{n} \sum_{i=1}^{n} g_\theta(X_i) - K(p_\theta, p_{\theta_0}) \right| = O_{\mathbf{P}}(n^{-1/2}).$$

If the same is true for the sequence $\{\hat{\theta}_n\}$:

$$\left| \frac{1}{n} \sum_{i=1}^{n} g_{\hat{\theta}_n}(X_i) - K(p_{\hat{\theta}_n}, p_{\theta_0}) \right| = O_{\mathbf{P}}(n^{-1/2}),$$

then (1.12), together with Lemma 1.3, would imply that $h(p_{\hat{\theta}_n}, p_{\theta_0}) = O_{\mathbf{P}}(n^{-1/4})$. In fact, one may expect that $\sigma_\theta \to 0$ as $h(p_\theta, p_{\theta_0}) \to 0$, and that in view of (1.14), $\sum_{i=1}^{n}(g_{\hat{\theta}_n}(X_i) - K(p_{\hat{\theta}_n}, p_{\theta_0}))/n$ converges with a rate faster than $O_{\mathbf{P}}(n^{-1/2})$. This would give $h(\hat{p}_n, p_0) = o_{\mathbf{P}}(n^{-1/4})$. We shall see that the rate of convergence ranges from $O_{\mathbf{P}}(n^{-1/2})$ for regular parametric models, to $o_{\mathbf{P}}(n^{-1/4})$ for moderately complex infinite-dimensional models. If the parameter space is even richer (has very large *entropy*), the rate can be even slower, or one may not have consistency at all. However, the power 1/4 appears to be something like a critical point.

Example 1.3. Estimating the mean in the binary choice model Here is another illustration of the application of extensions of the classical central limit theorem. Let us return to the situation of Example 1.1, Case (ii). There, we have

$$P(Y = 1 \mid Z = z) = F_0(z),$$

with F_0 an unknown increasing function satisfying $0 \leq F_0 \leq 1$. Suppose that we want to estimate the average probability of having a job. Write this as $\theta_0 = \theta_{F_0} = EY$. A good candidate for estimating θ_0 is of course $\bar{Y} = \sum_{i=1}^{n} Y_i/n$, the observed proportion of individuals with a job. We have

$$\sqrt{n}(\bar{Y} - \theta_0) \to^{\mathscr{L}} \mathscr{N}(0, \mathrm{var}(Y)),$$

with

$$\mathrm{var}(Y) = \int F_0(z)(1 - F_0(z))\, dQ(z) + \mathrm{var}(F_0(Z)). \tag{1.15}$$

Let $\hat{F}_n$ be as before the maximum likelihood estimator of F_0. Then $\sum_{i=1}^{n} \hat{F}_n(Z_i)/n$ is also a good candidate for estimating θ_0, but it turns out that this estimator is just $\bar{Y}$. Moreover, if the distribution Q of Z is completely unknown, then it is also the maximum likelihood estimator of θ_0. We shall not prove these two statements here (see Section 11.2 for more details). Instead, let us see what we can gain if we assume Q to be known. Since

$$\theta_0 = \int F_0(z)\, dQ(z),$$

the maximum likelihood estimator of θ_0 is then

$$\hat{\theta}_n = \theta_{\hat{F}_n} = \int \hat{F}_n(z)\, dQ(z).$$

More generally, we define

$$\theta_F = \int F(z)\, dQ(z).$$

Then for each $F \in \Lambda = \{F : \mathbf{R} \to [0,1],\ F \text{ increasing}\}$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(F(Z_i) - \theta_F\big) \to^{\mathscr{L}} \mathscr{N}\big(0, \text{var}(F(Z))\big).$$

Now, suppose that the process $\{\sum_{i=1}^{n}(F(Z_i) - \theta_F)/\sqrt{n} :\ F \in \Lambda\}$ is asymptotically continuous at F_0, in the sense that

$$(1.16) \qquad \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\hat{F}_n(Z_i) - \hat{\theta}_n\big) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(F_0(Z_i) - \theta_0\big) + o_{\mathbf{P}}(1).$$

This assumption can be seen as an extension of the central limit theorem (1.2).

Lemma 1.4 *Under (1.16),*

$$(1.17) \qquad \sqrt{n}(\hat{\theta}_n - \theta_0) \to^{\mathscr{L}} \mathscr{N}\left(0, \int F_0(z)(1 - F_0(z))\, dQ(z)\right).$$

Proof Since $\frac{1}{n}\sum_{i=1}^{n} \hat{F}_n(Z_i) = \bar{Y}$, we have

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n - \theta_0) &= \sqrt{n}(\bar{Y} - \theta_0) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\hat{F}_n(Z_i) - \hat{\theta}_n\big) \\
&= \sqrt{n}(\bar{Y} - \theta_0) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(F_0(Z_i) - \theta_0\big) + o_{\mathbf{P}}(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(Y_i - F_0(Z_i)\big) + o_{\mathbf{P}}(1).
\end{aligned}$$

The result now follows from the classical central limit theorem. □

So the asymptotic distribution of $\hat{\theta}_n$ has smaller variance than $\bar{Y}$. This is what we gain from knowing the distribution of Z.

1.2. Problems and complements

1.1. Verify (1.6) and (1.8). In fact, recall that if $\{p_\theta :\ \theta \in \Theta \subset \mathbf{R}\}$ is any sufficiently regular class, then for $l_\theta = d \log p_\theta / d\theta$ and $l'_\theta = dl_\theta/d\theta$, one has $El^2_{\theta_0}(X) = -El'_{\theta_0}(X) = I_{\theta_0}$. The quantity I_{θ_0} is called the Fisher information.