Problem set 4

Math 212a

November 1, 2001, Due Nov 13

Contents

1 Three standard counterexamples.

1.1 Non-dominated convergence.

Let $f_n := n\mathbf{1}_{(0,\frac{1}{n}]}$. This was one of the examples we studied in class where $\int_{\mathbf{R}} f_n dx \equiv 1$ while $\lim f_n(x) = 0$ for all x. So we get strict inequality in Fatou's lemma, and can not move the limit past the integral sign. So we know in advance that this can not be a case of dominated convergence.

1. Let g be defined by $g(x) := \sup_k f_k(x)$. We know that $g \notin \mathcal{L}_1$. But compute g and show explicitly that its integral diverges.

1.2 Order of integration matters for non-integrable functions.

2. Take $X = Y = [0,1]$ with standard Lebesgue measure. Let

$$f(x,y) = \begin{cases} \frac{1}{x^2} & \text{if} \quad 0 < y < x < 1 \\ -\frac{1}{y^2} & \text{if} \ \ 0 < x < y < 1 \\ 0 & \qquad\qquad \text{otherwise} \end{cases} .$$

Compute $\int_X \left(\int_Y f(x,y)dy\right) dx$ and $\int_Y \left(\int_X f(x,y)dx\right) dy$ and verify that they are not equal. This implies that $f \notin \mathcal{L}_1(X \times Y, \mathbf{R})$ for otherwise Fubini would apply. Show directly that both f^+ and f^- have infinite integrals on $X \times Y$.

1.3 Order of integration matters in the non-σ-finite case.

Let $X = Y = [0,1]$ as before. Take $\mathcal{F}$ to be the usual Lebesgue measurable sets and m to be Lebesgue measure. Take $\mathcal{G}$ to be "all" subsets, and take n to be

$$n(A) = \text{number of elements in} \ \ A$$

(in particular $n(A) = \infty$ if A is not a finite set). Since $[0,1]$ is not the countable union of finite sets, it is not σ-finite. So we expect that Fubini might fail. Let $\Delta \subset X \times Y$ be the diagonal, i.e.

$$\Delta := \{(x,y)|x = y\}.$$

3. Show that $\Delta \in \mathcal{F} \times \mathcal{G}$. Let

$$f := \mathbf{1}_\Delta.$$

Evaluate $\int_X \left(\int_Y f(x,y)dy\right) dx$ and $\int_Y \left(\int_X f(x,y)dx\right) dy$ and verify that they are not equal.

2 L^1 is not reflexive.

If $1 < p < \infty$ and

$$\frac{1}{p} + \frac{1}{q} = 1.$$

We proved that

$$(L^p)^* = L^q$$

and therefore

$$((L^p)^*)^* = L^p.$$

For any Banach space B, the space B^* is a again a Banach space, and there is a natural embedding

$$B \to (B^*)^*$$

given by evaluation: an element $f \in B$ goes into the linear function on B^* sending

$$\ell \mapsto \ell(f) \quad \ell \in B^*.$$

This map has norm ≤ 1 by its very definition, and the Hahn-Banach theorem guarantees that we can find an $\ell \in B^*$ such that $|\ell(f)| = \|f\|$ so this map $B \to (B^*)^*$ an isometry, i.e. is norm-preserving, and hence injective.

A Banach space B is called **reflexive** if this map is is surjective, i.e. and isomorphism. So our result above about L^p and L^q when $1 < p < \infty$ says that the spaces L^p are reflexive when $1 < p < \infty$.

We know that $(L^1)^* = L^\infty$. The next few exercises will explain why $(L^\infty)^*$ is strictly bigger than L^1 when we are dealing (say) with Lebesgue measure on the unit interval $[0, 1]$.

Recall that the norm $\|\cdot\|$ is the essential supremum norm as defined in class on the space of essentially bounded Lebesgue measurable functions (now on the interval $[0, 1]$). An element f of $L^\infty = L^\infty([0, 1])$ is said to be **essentially continuous** if there is a continuous (and hence bounded) function on $[0, 1]$ which agrees with f except on a set of measure zero. So if we have identified two functions which differ on a set of measure zero, then we can say that the set of essentially continuous functions is the image of the space of continuous functions under the (norm preserving) map which maps each function to its equivalence class. Denote the space of essentially continuous functions by $C_{ess} = C_{ess}([0, 1])$. For example, Dirichlet's function which is zero on the rationals and one on the irrationals is essentially continuous, since it is equivalent to the constant one. Notice that C_{ess} is a closed subspace of L^∞ since the uniform limit of a sequence of continuous functions is continuous.

4. Let $\ell \in (L^\infty)^*$. Show that there is a $g \in L^1$ such that

$$\ell(f) = \int_{[0,1]} fg dx \quad \forall\, f \in C_{ess},$$

and that g is unique (up to equivalence).

So the the guts of the issue is

5. a) Show that $C_{ess} \neq L^\infty$. That is, show that there is a bounded Lebesgue measurable function on $[0, 1]$ which is not essentially continuous. b) Show that there is a non-zero continuous linear functional on L^∞ which vanishes on C_{ess}. [HInt: a) As part of problem 4 of the first problem set, you constructed an everywhere dense open subset of $[0, 1]$ whose measure is strictly less than 1.]

3 The Central Limit Theorem.

The purpose of the rest of this problem set is to work through the proof of the "central limit theorem" of probability theory. Roughly speaking, this theorem asserts that if the random variable S_n is the sum of many independent random variables

$$S_n = X_1 + \cdots + X_n$$

all with mean zero and finite variance then under appropriate additional hypotheses

$$\frac{1}{s_n} S_n$$

is approximately normally distributed where

$$s_n^2 := \text{var}(S_n) = \sigma_1^2 + \cdots + +\sigma_n^2, \quad \sigma_i^2 := \text{var}(X_i).$$

The actual condition, equation (12) below, is rather technical. But roughly speaking it says that no one X_i outweighs the others. The condition of mean zero for each of the X_i is just a matter of convenience in formulation. Otherwise we would replace the statement by the the assertion that $\frac{1}{s_n}(S_n - E(S_n))$ is approximately normally distributed. The phrase "approximately normally distributed" will be taken to mean in the sense of weak convergence: For each bounded continuous function f we have

$$\lim_{n\to\infty} E\left(f(\frac{1}{s_n} S_n)\right) \to E(f(N)) \tag{1}$$

where N is a unit normal random variable:

$$P(N \in [a, b]) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

It is enough to prove this for a set of functions f which are dense in the uniform norm. In fact, the proof outlined below will assume that f is three times differentiable with a uniform bound on its first three derivatives.

The idea of the proof presented here is the same as the idea behind the proof that we gave in Problem Set 2 of Poisson's law of small numbers. We replace each X_k successively in the sum giving S_n by a normal random variable Z_k having mean zero and variance σ_k. We then estimate the total error committed by these substitutions.

Philosophically, the import of the theorem is that if we observe an effect whose variation is the sum of many small positive and negative random contributions, no one outweighing the others, then the effect will be distributed according to a normal distribution with a finite mean and variance. So this theorem accounts for the ubiquity of the normal distribution in science.

3.1 Push Forward.

Let $(X, \mathcal{F}, m)$ be a measure space and $(Y, \mathcal{G})$ is a space with a σ-field. We called a map $f : X \to Y$ " measurable" if $f^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{G}$ and then defined the push forward $f_* m$ of the measure m by

$$f_*(B) = m(f^{-1}(B)).$$

We will need some formulas for pushforward. In the cases we consider we will have $X = [0,1]$ (so frequently we will think of push forward as the process of "simulating" a random variable using a random number generator as explained in Problem set 2), or we will have $X = \mathbf{R}^n$ or some nice subset of $\mathbf{R}^n$. Similarly for Y.

In all our examples, measures will either be discrete or have densities relative to Lebesgue measure. A discrete measure ν integrates functions by the formula

$$\langle \nu, \phi \rangle = \int \phi \nu = \sum \phi(x_k) r_k, \quad r_k = \nu(\{x_k\}).$$

The push forward of the discrete measure ν by a map f is concentrated on the set

$$\{y_\ell\} = \{f(x_k)\} \ \text{ with } \ (f_*\nu)(\{y_\ell\}) = \sum_{f(x_k) = y_\ell} \nu(\{x_k\}).$$

For functions we have

$$\langle f_*\nu, \phi \rangle = \langle \nu, f^*\phi \rangle.$$

In this equation and in what follows we use the "pull back" notation

$$f^*\phi := \phi \circ f.$$

The measures with density have

$$\langle \nu, \phi \rangle = \int \phi(x)\rho(x) dx$$

where ρ is the density.

For the standard Lebesgue linear measure du (so density one) and for $f : u \mapsto x$ a map of real intervals which is one to one, differentiable and with differentiable inverse we have

$$f_* du = |f'(u(x))|^{-1} dx = |du/dx| dx. \tag{2}$$

Proof. By the change of variables formula

$$\int \psi dx = \int f^*\psi |dx/du| du.$$

Set

$$\psi = \phi |dx/du|^{-1}.$$

The change of variables formula becomes

$$\begin{aligned}\int \phi |dx/du|^{-1} dx &= \int f^* \phi f^* |dx/du|^{-1} |dx/du| du \\ &= \int f^* \phi du\end{aligned}$$

by the chain rule since

$$\frac{dx}{du}(x(u))\frac{dx}{du}(u) \equiv 1.$$

QED

Example:

$$x = -(1/\lambda)\ln u.$$

This maps the interval $(0,1]$ onto the positive numbers (reversing orientation so 1 goes to 0 and 0 goes to ∞). So

$$u = e^{-\lambda x} \quad \text{and} \quad |du/dx| = \lambda e^{-\lambda x}.$$

The probability law on $\mathbf{R}^+$ with density $\lambda e^{-\lambda x}$ is known as the exponential law. So the function $-(1/\lambda)\ln u$ simulates the exponential law. The MATLAB command

$-(1/\lambda) * \log(\text{rand(M,N)})$

will produce an M×N matrix whose entries are non-negative numbers independently exponentially distributed with parameter λ.

Example:

$$r = \sqrt{-2\ln u} \quad \text{so} \quad u = e^{-r^2/2}.$$

The push forward of the uniform measure du is

$$e^{-r^2/2} r dr.$$

The same argument shows that the push forward of $\rho(u)du$ is given by

$$f_*[\rho du] = \rho(u(x))|du/dx|dx. \tag{3}$$

The formula (3) works in n dimensions where du/dx is interpreted as the Jacobian matrix.

Example. The area in polar coordinates in the plane is given by

$$r dr d\theta.$$

If we set $S = r^2$ we can write this as

$$dA = \frac{1}{2} dS d\theta.$$

So, for example, if we want to describe a uniform probability measure on the unit disk, it will be given by

$$\frac{1}{2\pi} dS d\theta.$$

Notice that in this formula S is uniformly distributed on $[0, 1]$ and θ is uniformly distributed on $[0, 2\pi]$ and they are independent. Suppose we start with S and θ and (changing from our previous notation) set

$$r = \sqrt{-2 \ln S}.$$

We get the measure

$$\frac{1}{2\pi} e^{-r^2/2} r dr d\theta$$

for a probability measure on the plane. In particular the total integral of the above expression is one.

If we change from polar coordinates to rectangular coordinates:

$$x = r \cos\theta, \quad y = r \sin\theta$$

the density becomes

$$\frac{1}{2\pi} e^{-(x^2+y^2)/2} dx dy.$$

Notice that the random variables X and Y (projections onto the coordinate axes) are independent, each with density

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

In particular

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1.$$

Example. The rejection method.

Suppose that V_1 and V_2 are independent random variables uniformly distributed on $[-1, 1]$ and that

$$S = V_1^2 + V_2^2.$$

and so

$$\cos\theta = \frac{V_1}{S^{1/2}}, \quad \sin\theta = \frac{V_2}{S^{1/2}}$$

if we use V_1, V_2 as rectangular coordinates.

The vector (V_1, V_2) is uniformly distributed over the square $[-1, 1] \times [-1, 1]$. We would like to have a measure uniformly distributed over the unit disk. So add an instruction which rejects the output and tries again if $S > 1$. Among the accepted outputs, the probability of ending up in any region of the disk is proportional to the area of the region. Thus we have produced a uniform distribution on the unit disk, and we know from the preceding discussion that S, the radius squared, and the angle θ are independent and uniformly distributed, S on $[0,1]$ and θ on $[0,2\pi]$.

Substituting into the above we get that

$$X = (-2\ln S)^{1/2}\frac{V_1}{S^{1/2}}, \quad Y = (-2\ln S)^{1/2}\frac{V_2}{S^{1/2}}$$

are independent normally distributed random variables. So we can simulate two independent normally distributed random variables (X,Y) by the following algorithm

- Generate two independent random numbers U_1 and U_2 using the random number generator. (In MATLAB $[U_1.U_2] =$ rand(1,2)).
- Set
$$V_1 = 2U_1 - 1,\ V_2 = 2U_2 - 1,\ S = V_1^2 + V_2^2.$$
- If $S > 1$ return to step 1. Otherwise
- Set

$$X = V_1(-2\ln S/S)^{1/2}, \quad Y = V_2(-2\ln S/S)^{1/2}.$$

Of course, MATLAB has its own built in normal random number generator: the command randn(M,N) produces an M×N matrix whose entries are independent normally distributed random variables.

Push forward from higher to lower dimension involves summation or integration. (multiple integral as iterated integral). These may or may not converge. But they will converge if the total measure is finite, for example for a probability measure.

Example. $(x,y) \mapsto x+y$ is a map $f : \mathbf{R}^2 \to \mathbf{R}$. If m is a density on $\mathbf{R}^2$ and ϕ is a function on $\mathbf{R}$, then

$$\int\int f^*\phi m(x,y)dxdy = \int\int \phi(x+y)m(x,y)dxdy.$$

Set $z = x + y$ and $u = y$. Apply the change of variables formula; the Jacobian is identically 1 . So the integral becomes

$$\int\int \phi(z)m(z-u,u)dzdu = \int \phi(z)\int m(z-u,u)dudz$$

by iterated integration. So

$$f^*m = \rho$$

where

$$\rho(z) = \int m(z-u,u)du.$$

An important special case is where $m(x, y) = r(x)s(y)$ (independence). Then

$$\rho(z) = \int r(z-u)s(u)du$$

is called the convolution of r and s and written as

$$\rho = r \star s.$$

Notice that

$$r \star s = s \star r.$$

This differs slightly from the convention we used earlier in the course which involves a factor of $1/\sqrt{2\pi}$. But this is the standard probabilists' convention. (The standard probabilists' convention about the Fourier transform differs from the one we have been using both by a factor of $\sqrt{2\pi}$ and by the sign in the exponential.)

3.2 Moment generating functions

For any random variable X try to define

$$M_X(t) := E(e^{tX}).$$

Of course this may only be defined for a limited range of t due to convergence problems. For example, suppose that X is exponentially distributed with parameter λ. Then

$$M_X(t) = \lambda \int_0^\infty e^{(t-\lambda)x}dx$$

diverges for $t \geq \lambda$. But for $t < \lambda$ the integral converges to

$$\frac{\lambda}{\lambda - t}.$$

For any random variable, if t is interior to the convergence domain of the moment generating function, M_X, then M_X is differentiable to all orders at t and we have

$$M^{(n)}(t) = E(X^n e^{tX}).$$

In particular, if 0 is interior to the domain of convergence, we have

$$E(X^n) = M^{(n)}(0).$$

For example, for the exponential distribution the $n-$th derivative of the moment generating function is given by

$$\frac{n!\lambda}{(\lambda - t)^{n+1}}$$

and so

$$E(X^n) = n!\lambda^{-n}.$$

So

$$E(X)=\frac{1}{\lambda},\quad E(X^2)=\frac{2}{\lambda^2}$$

and so

$$\mathrm{Var}(X)=\frac{1}{\lambda^2}.$$

In statistical mechanics the tradition is to study $M(-\beta)=P(\beta)$ which is called the *partition function.*

Suppose we set

$$X=\sigma N+m. \tag{4}$$

Then

$$E(X)=m \quad \text{and } \mathrm{Var}(X)=\sigma^2 \tag{5}$$

while the change of variables formula implies that the density of X is given by

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}. \tag{6}$$

Conversely, suppose that Z is a random variable with density

$$Ce^{-q(x)/2}$$

where

$$q(x)=ax^2+2bx+c,\quad a>0,$$

is a quadratic polynomial, and C is a constant. Completing the square allows us to write

$$q(x)=a(x+\frac{b}{a})^2-\frac{b^2}{a^2}+c$$

and the fact that the total integral must be one constrains the constants so that

$$Ce^{c-b^2/a^2}=\frac{1}{\sigma\sqrt{2\pi}}\quad\text{where}\quad\sigma=\frac{1}{\sqrt{a}}.$$

Hence Z is equivalent in law to a random variable of the form (4) where

$$\sigma=\frac{1}{\sqrt{a}},\quad m=-\frac{b}{a}.$$

A random variable whose density is proportional to $e^{-q(x)/2}$ for some quadratic polynomial is called a *Gaussian* random variable. We have seen that such a random variable is completely determined in law by its mean and variance, and is equivalent in law to a random variable of the form (4).

In particular, if X is a Gaussian with mean m_X and variance σ_X^2 and Y is a Gaussian with mean m_Y and variance σ_Y^2, and if X and Y are independent, then $X+Y$ is a Gaussian with mean m_X+m_Y and variance $\sigma_X^2+\sigma_Y^2$.

3.3 Central limit theorem by the coupling method.

First a preliminary technical lemma:

Lemma 1 *Let Z be a Gaussian random variable with mean zero and variance σ^2. Then there exists a constant L, independent of σ, s or ϵ such that*

$$E_{|Z|>\epsilon s}(Z^2) \le L\frac{\sigma^3}{\epsilon s}. \tag{7}$$

We use the notation $E_A(Y)$ to mean the expected value of Y over the set A, i.e. $E(\mathbf{1}_A(Y)Y)$. For example the left hand side of the inequality in the lemma

$$2\frac{1}{\sigma\sqrt{2\pi}}\int_{\epsilon s}^{\infty} z^2 e^{\frac{z^2}{2\sigma^2}}dz.$$

6. Prove the lemma. Hint: Make the change of variables $z = \sigma x$. Let $a = \frac{\epsilon s}{\sigma}$. Show that $a\int_a^\infty x^2 e^{-x^2/2}dx$ is bounded.]

Let $X_1, X_2, X_3, \ldots, X_n, \ldots$ be independent random variables with means zero and with variances σ_i^2. Let

$$S_n = X_1 + \cdots + X_n$$

so that

$$s_n^2 \stackrel{def}{=} \mathrm{Var}S_n = \sigma_1^2 + \cdots \sigma_n^2.$$

We wish to prove that under condition (12) to be stated below we have

$$\lim_{n\to\infty} E\left(f\left(\frac{1}{s_n}S_n\right)\right) \to E(f(N)) \tag{8}$$

for any three times differentiable function, f with bounded third derivatives.

The method of proof is to choose Gaussian random variables

$$Z_1, Z_2, \ldots, Z_n, \ldots$$

all independent of the X_i's and each other, with

$$\mathrm{Var}(Z_i) = \sigma_i^2$$

and make the successive substitutions

$$\begin{aligned} S_n = X_1 + \cdots + X_{n-1} + X_n &\to X_1 + \cdots + X_{n-1} + Z_n \\ &\to X_1 + \cdots + Z_{n-1} + Z_n \\ &\vdots \\ &\to Z_1 + Z_2 + \cdots + Z_{n-1} + Z_n \stackrel{def}{=} Z, \end{aligned}$$

and to estimate the difference in expectation at each stage of substitution. At the end of the substitutions, Z is a Gaussian with variance s_n^2 and hence $(1/s_n)Z$ is equal in law to the (unit) normal distribution, N.

The difference at the $(n-k)-$th stage is

$$E(f(\frac{1}{s_n}(T_k + X_k))) - E(f(\frac{1}{s_n}(T_k + Z_k)))$$

where

$$T_k = X_1 + \cdots + X_{k-1} + Z_{k+1} + \cdots + Z_n$$

is the sum of all the random variables which are unchanged during this particular substitution. Now we certainly have

$$|E(f(\frac{1}{s_n}(T_k+X_k)) - E(\frac{1}{s_n}(T_k+Z_k))| \leq \sup_u |E(f(u+\frac{1}{s_n}X_k) - E(f(u+\frac{1}{s_n}Z_k))|$$

since, if we set

$$h(u) = E(f(u + \frac{1}{s_n}X_k)) - E(f(u + \frac{1}{s_n}Z_k))$$

we have

$$E(h(\frac{1}{s_n}T_k)) = E(f(\frac{1}{s_n}(T_k + X_k))) - E(f(\frac{1}{s_n}(T_k + Z_k))).$$

So it will be enough for us to get a bound on

$$|E(f(u + \frac{1}{s_n}X_k)) - E(f(u + \frac{1}{s_n}Z_k))|,$$

valid for all u.

Consider the Taylor expansion, with remainder, of f about the point u:

$$f(u + y) = f(u) + yf'(u) + \frac{1}{2}f''(u)y^2 + g(u, y),$$

where

$$g(u, y) = f(u + y) - [f(u) + yf'(u) + \frac{1}{2}f''(u)y^2] \tag{9}$$

is of the form

$$g(u, y) = \frac{1}{3!}f'''(u^*)y^3 \tag{10}$$

where u^* is some point between u and $u + y$. Since $f(u)$ is a constant as far as y is concerned, taking its expectations either with respect to Z_k or X_k gives the same value, $f(u)$. The expectations $E(X_k)$ and $E(Z_k)$ both vanish, and the expectations

$$E(X_k^2) = E(Z_k^2).$$

So

$$E(f(u)+\frac{1}{s_n}X_k f'(u)+\frac{1}{2}f''(u)[\frac{1}{s_n}X_k]^2) = E(f(u)+\frac{1}{s_n}Z_k f'(u)+\frac{1}{2}f''(u)[\frac{1}{s_n}Z_k]^2)$$

and hence

$$E(f(u+\frac{1}{s_n}X_k) - E(f(u+\frac{1}{s_n}Z_k)) = E(g(u,\frac{1}{s_n}X_k) - E(g(u,\frac{1}{s_n}Z_k))$$

so

$$|E(f(u+\frac{1}{s_n}X_k)) - E(f(u+\frac{1}{s_n}Z_k))| \le E(|g(u,\frac{1}{s_n}(X_k)|) + E(|g(u,\frac{1}{s_n}Z_k)|)$$

and we shall estimate each of the terms on the right separately. From (10) we have

$$|g(u,y)| \le K_1 y^3, \quad K_1 = \frac{1}{3!}\sup_u |f'''(u)|$$

valid for all y while for large $|y|$ the definition (9) implies that

$$|g(u,y)| \le K_2|y|^2$$

where K_2 can be expressed in terms of $\sup|f|$, $\sup|f'|$, and $\sup|f''|$. So there is a constant K such that

$$|g(u,y)| \le K\min(|y|^2,|y|^3)$$

for all u and y. Of course we want to use the $|y|^3$ estimate for small $|y|$ and the $|y|^2$ estimate for large $|y|$.

In any event, given $\epsilon > 0$ we have

$$\begin{aligned}
E(|g(u,\frac{1}{s_n}(X_k)|) &= E_{|X_k|\le\epsilon s_k}(|g(u,\frac{1}{s_n}X_k)|) + E_{|X_k|>\epsilon s_k}(|g(u,\frac{1}{s_n}X_k)|) \\
&\le KE_{|X_k|\le\epsilon s_k}(|X_k|^3/s_n^3) + KE_{|X_k|>\epsilon s_n}(X_k^2/s_n^2) \\
&\le K\epsilon E_{|X_k|\le\epsilon s_k}(|X_k|^2/s_n^2) + KE_{|X_k|>\epsilon s_n}(X_k^2/s_n^2) \\
&\le K\frac{\epsilon}{s_n^2}E(|X_k|^2) + KE_{|X_k|>\epsilon s_n}(X_k^2/s_n^2) \\
&= K\epsilon\frac{\sigma_k^2}{s_n^2} + \frac{K}{s_n^2}E_{|X_k|>\epsilon s_n}(X_k^2).
\end{aligned}$$

We can make a similar estimate with Z_k instead of X_k to obtain

$$\begin{aligned}
E(|g(u,\frac{1}{s_n}Z_k)|) &\le K\epsilon\frac{\sigma_k^2}{s_n^2} + \frac{K}{s_n^2}E_{|Z_k|>\epsilon s_n}(Z_k^2) \\
&\le K\epsilon\frac{\sigma_k^2}{s_n^2} + \left(\frac{KL}{\epsilon}\right)\left(\frac{\sigma_k^3}{s_n^3}\right)
\end{aligned}$$

where, in passing from the first line to the second we made use of (7). We now add all these inequalities over k, using the facts that

$$\frac{1}{s_n}(Z_1 + \cdots + Z_n) \overset{law}{=} N$$

and

$$\sigma_1^2 + \cdots + \cdots \sigma_n^2 = s_n^2$$

to obtain

$$|E(f(\frac{1}{s_n}S_n)) - E(f(N))| \leq 2K\epsilon + \frac{K}{s_n^2}\sum_k E_{|X_k|>\epsilon s_n}(X_k^2) + \frac{KL}{\epsilon}\sum \frac{\sigma_k^3}{s_n^3}. \quad (11)$$

We can now state the central limit theorem:

Theorem 1 *Suppose that for any* $\epsilon > 0$

$$\lim_{n\to\infty}\left(\frac{1}{s_n^2}\right)\sum_k E_{|X_k|>\epsilon s_n}(X_k^2) = 0. \quad (12)$$

Then for any three times differentiable function, f, which is bounded together with its first three derivatives in absolute value we have

$$E\left(f\left(\frac{1}{s_n}S_n\right)\right) \to E(f(N)). \quad (13)$$

where N is the normally distributed random variable.

Proof. We are assuming that the second term in (11) goes to zero with n for any $\epsilon > 0$. So the question is estimating the third term. The sum giving the third term is maximized by $\max_k \sigma_k/s_n$ since the sum with squares instead of cubes adds up to 1.

7. Show that (12) implies that $\max_k \sigma_k/s_n \to 0$. [Hint: Break the expression $E\left(\frac{X_k^2}{s_n^2}\right)$ up into two parts, one $E_{|X_k|\leq\delta s_n}$ and the other $E_{|X_k|>\delta s_n}$. Show that by choosing $\delta < \epsilon^2$ the right hand side of (11) can be estimated by some multiple of ϵ.]

A loose interpretation of the condition (12) is that it is a little stronger than requiring that no one variance outweighs the others in the sense that

$$\frac{\sigma_k}{s_n} \to 0.$$

8. Suppose that all the X_ks are identically distributed. This means that they are all equal in law to a common X with variance, say $\sigma > 0$. In this case

$$s_n^2 = n\sigma^2.$$

Show that (12) is satisfied in this case.