ASYMPTOTIC METHODS AND STATISTICAL APPLICATIONS[1]

Heather Battey

Department of Mathematics, Imperial College London

March 14, 2020

[1]Thanks are due to Daniel Garcia Rasines and Scott Ward for indicating an error in an earlier version of these notes. Section 1.3 and much of section 1.2 of Chapter 1 are slight adaptations of notes by Prof. X. Wu, Department of Mathematics, Imperial College London.

Preface

Many problems of practical interest do not possess an exact analytic solution, or have solutions too complicated to be useful. The present course develops techniques for obtaining approximate analytic solutions, where the approximation error becomes small as a key parameter, α say, tends to a limit α_0. Such approximations allow, for instance, quantification of the performance of statistical procedures in terms of intrinsic features or key tuning parameters.

While approximate solutions are obtained by letting $\alpha \to \alpha_0$, it is important to remember that this limiting operation is purely notional and without direct physical significance. The expression so obtained is an asymptotic form rather than a limit, i.e., it depends on α, and the objective is to derive good approximations for the particular α involved in an application. For instance, in many statistical applications, the limiting operation is $n \to \infty$ where n is the sample size. The central limit theorem provides a very good approximation to a suitably scaled sum of random variables under idealized conditions if n is very large. Approximations that work well even for small n are to be preferred and the adequacy of the asymptotic approximation always needs consideration. See Cox (1988) for an enlarged discussion of the role of asymptotic theory in statistics.

The content of this course was chosen with a statistical audience in mind and the examples used are primarily from statistics. However many of the ideas, being mathematical in basis, are more widely applicable.

Contents

ASYMPTOTIC METHODS AND STATISTICAL APPLICATIONS

Chapter 1: Preliminaries

Heather Battey | February 14, 2020

1.1 Some crude asymptotic theory for statistics

1.1.1 Orders of magnitude

Let f_n be a positive nonstochastic decreasing sequence, e.g., $\{1/n\}_{n\in\mathbb{N}}$, $\{1/\sqrt{n}\}_{n\in\mathbb{N}}$. We write $T_n = O(f_n)$ if T_n/f_n is bounded for all sufficiently large n. We write $T_n = o(f_n)$ if $T_n/f_n \to 0$ as $n \to \infty$. With this definition it is clear that if $T_n/f_n \to c < \infty$ and f_n is a decreasing sequence, then T_n converges to zero. Thus we may write both $T_n = o(1)$ and $T_n = O(f_n)$, but the latter provides more information.

1.1.2 Stochastic orders of magnitude

Definition 1.1.1. A sequence of random variables $\{T_n\}_{n\in\mathbb{N}}$ converges in probability to a random variable (or a constant) T if, for all $\varepsilon > 0$, $\mathrm{pr}(|T_n - T| > \varepsilon) \to 0$ as $n \to \infty$. Equivalently, for all $\varepsilon > 0$, $\delta > 0$ there exists a $n_0 = n_0(\varepsilon, \delta)$ such that for all $n \geq n_0$, $\mathrm{pr}(|T_n - T| > \varepsilon) < \delta$.

Remark 1.1.2. The above definition can be adapted for vector, matrix or function valued random sequences $\{T_n\}_{n\in\mathbb{N}}$ by replacing the absolute value by a suitable norm.

Remark 1.1.3. For manipulating statements involving probabilities of events, it is sometimes helpful think in the following terms. There are basic random variables X, measurable mappings from a sample space $\mathcal{S}$ to a measurable space $\mathcal{X}$. Often $\mathcal{X} = \mathbb{R}^p$ in applications of interest but in principle $\mathcal{X}$ could be any separable space. The statistic T_n is a (measurable) function $T_n = T_n(X_1, \ldots, X_n)$ which is itself a random variable, $T_n : \Omega \to \mathcal{T}$, where the sample space Ω is, for $\{X_i\}_{i=1}^n$ independent and identically distributed (i.i.d.) copies of X, the product space $\mathcal{S}^n$. Fixing an element $\omega \in \Omega$ gives a *realization* $t_n = T_n(\omega)$ of T_n. Upper-case letters are typically used for random variables and lower-case for their realizations. Thus, when we refer to an event such as

$\{|T_n| > \varepsilon\}$ we are specifying a subset $\{\omega \in \Omega : |T_n(\omega)| > \varepsilon\} \subseteq \Omega$.

Let f_n be a positive nonstochastic decreasing sequence. We write $T_n = O_p(f_n)$ if for all $\delta > 0$, there exists a $C < \infty$ and $n_0 > 0$ such that $\mathrm{pr}(|T_n| > C f_n) < \delta$ for all $n > n_0$. We write $T_n = o_p(f_n)$ if $T_n/f_n \to_p 0$.

Example 1.1.4. Let $\{X_i\}_{i=1}^n$ be i.i.d. with mean μ and variance σ^2. Let $T_n = n^{-1}\sum_{i=1}^n X_i$. By Markov's inequality,

$$\mathrm{pr}(|T_n - \mu| > C) \leq \frac{\mathbb{E}(T_n - \mu)^2}{C^2} = \frac{\mathrm{var}(T_n)}{C^2} = \frac{\sigma^2}{nC^2}$$

Therefore any $C > \kappa n^{-1/2}$ with $\kappa = \sqrt{\sigma^2/\delta}$ ensures $\mathrm{pr}(|T_n - \mu| > C) < \delta$. In other words $(T_n - \mu) = O_p(n^{-1/2})$.

Various statements about stochastic orders of magnitude will now be proved. This will aid understanding of the concepts involved, and the proof strategies are applicable in a range of settings.

Proposition 1.1.5. For a finite constant c, $T_n \to_p c \Rightarrow T_n = O_p(1)$ and for $c = 0$, $T_n = o_p(1) \iff T_n \to_p 0$.

Proof. By the definition of convergence in probability to c, for all $\varepsilon, \delta > 0$ there exists an $n_0 = n_0(\varepsilon, \delta)$ such that, for all $n > n_0$, $\mathrm{pr}(|T_n - c| > \varepsilon) < \delta$. To show $T_n = O_p(1)$, we must show the existence of a $C < \infty$ and n_0 such that for $\delta > 0$ and for all $n > n_0$, $\mathrm{pr}(|T_n| > C) < \delta$. We have

$$\begin{aligned} \mathrm{pr}(|T_n| > C) &= \mathrm{pr}(|T_n - c + c| > C) \\ &\leq \mathrm{pr}(|T_n - c| + |c| > C) \\ &= \mathrm{pr}(|T_n - c| > C - |c|). \end{aligned}$$

The inequality follows because $|T_n - c + c| \leq |T_n - c| + |c|$ and the probability of the larger object being greater than C is higher than the probability of the smaller object being greater than C. Fixing an arbitrary $\varepsilon > 0$ and letting $C = |c| + \varepsilon$, then

$$\mathrm{pr}(|T_n| > C) \leq \mathrm{pr}(|T_n - c| > \varepsilon) < \delta$$

for all $\delta > 0$ by $T_n \to_p c$. $\square$

Remark 1.1.6. If $\mathrm{pr}(|T_n| > C) \to 0$ for *some* $C < \infty$, then $T_n = O_p(1)$, while if $\mathrm{pr}(|T_n| > C) \to 0$ for *all* $C > 0$, $T_n = o_p(1)$.

Proposition 1.1.7. Let f_n be a positive deterministic sequence.

(i) $T_n = o_p(f_n) \Rightarrow T_n = O_p(f_n)$.

(ii) $T_n = O_p(f_n) \Rightarrow T_n = o_p(g_n)$ for any positive deterministic sequence g_n such that $f_n/g_n \longrightarrow 0$.

(iii) $T_n = O_p\left((\mathbb{E}|T_n|^r)^{1/r}\right)$ for $r > 0$.

Proof. (i) The proof is that of proposition 1.1.5 with T_n replaced by T_n/f_n.

(ii) For any $\varepsilon > 0$ and for some $C < \infty$, there exists an $n_0(\varepsilon, C)$ such that $\varepsilon g_n/f_n > C$ for all $n > n_0(\varepsilon, C)$, hence

$$\mathrm{pr}(|T_n| > \varepsilon g_n) = \mathrm{pr}(|T_n| > \varepsilon f_n g_n/f_n) \leq \mathrm{pr}(|T_n| > C f_n).$$

(iii) By Markov's inequality, for any $\delta > 0$, there exists a $C = C(\delta) < \infty$ such that

$$\mathrm{pr}(|T_n| > C(\mathbb{E}|T_n|^r)^{1/r}) \leq \frac{\mathbb{E}|T_n|^r}{C^r\mathbb{E}|T_n|^r} = C^{-r} < \delta.$$

□

Proposition 1.1.8. [Algebra of stochastic orders] Let f_n and g_n be positive deterministic sequences.

1. If $T_n = O_p(f_n)$ and $S_n = O_p(g_n)$ then

   (i) $T_nS_n = O_p(f_ng_n)$.

   (ii) $T_n + S_n = O_p(\max\{f_n, g_n\})$.

2. If $T_n = O_p(f_n)$ and $S_n = o_p(g_n)$, then $T_nS_n = o_p(f_ng_n)$

Remark 1.1.9. We can replace "O" by "o" everywhere in 1.

Proof. 1. By $T_n = O_p(f_n)$ and $S_n = O_p(g_n)$ we know that, for any $\delta > 0$, there exist a $C, D < \infty$ and $n_1, n_2 > 0$ such that

$$\begin{aligned} \mathrm{pr}(|T_n| > Cf_n) &< \delta/2 \quad \forall n \geq n_1 \\ \mathrm{pr}(|S_n| > Dg_n) &< \delta/2 \quad \forall n \geq n_2. \end{aligned}$$

(i) For all $n \geq \max\{n_1, n_2\} > 0$, we have

$$
\begin{aligned}
& \operatorname{pr}(|T_n S_n| > CDf_n g_n) \\
= \; & \operatorname{pr}(|T_n||S_n| > Cf_n Dg_n) \\
= \; & \operatorname{pr}\Big(\Big\{\{|T_n||S_n| > Cf_n Dg_n\} \cap \{|S_n|/Dg_n > 1\}\Big\} \cup \Big\{\{|T_n||S_n| > Cf_n Dg_n\} \cap \{|S_n|/Dg_n \leq 1\}\Big\}\Big) \\
= \; & \operatorname{pr}\Big(\{|T_n||S_n| > Cf_n Dg_n\} \cap \{|S_n|/Dg_n > 1\}\Big) + \operatorname{pr}\Big(\{|T_n||S_n| > Cf_n Dg_n\} \cap \{|S_n|/Dg_n \leq 1\}\Big) \\
\leq \; & \operatorname{pr}(|T_n| > Cf_n) + \operatorname{pr}(|S_n| > Dg_n) < \delta.
\end{aligned}
$$

The final line follows because the probability of the joint event $\big\{|T_n||S_n| > Cf_n Dg_n\big\} \cap \big\{|S_n|/Dg_n > 1\big\}$ must be smaller than the probability of the single event $\{|T_n| > Cf_n\}$, whilst if events $A \triangleq \big\{|T_n||S_n| > Cf_n Dg_n\big\}$ and $B \triangleq \big\{|S_n|/Dg_n \leq 1\big\}$ both occur, a fortiori (replacing $|S_n|/Dg_n$ by 1), event $E \triangleq \big\{|T_n| > Cf_n\big\}$ occurs, i.e. $(A \cap B) \subseteq E$, thus $\operatorname{pr}(A \cap B) \leq \operatorname{pr}(E)$.

(ii) For all $n \geq \max\{n_1, n_2\} > 0$, we have

$$
\begin{aligned}
& \operatorname{pr}(|T_n + S_n| > (C + D)\max\{f_n, g_n\}) \\
\leq \; & \operatorname{pr}(|T_n| + |S_n| > (C + D)\max\{f_n, g_n\}) \\
& \text{(by the triangle inequality)} \\
= \; & \operatorname{pr}\Big(\{|T_n| + |S_n| > (C + D)\max\{f_n, g_n\}\} \cap \{|S_n| > D\max\{f_n, g_n\}\}\Big) \\
& \quad + \operatorname{pr}\Big(\{|T_n| + |S_n| > (C + D)\max\{f_n, g_n\}\} \cap \{|S_n| \leq D\max\{f_n, g_n\}\}\Big) \\
\leq \; & \operatorname{pr}(|S_n| > Dg_n) + \operatorname{pr}(|T_n| > Cf_n).
\end{aligned}
$$

The last line comes from the following argument. The probability of the joint event

$$\{\{|T_n| + |S_n| > (C + D)\max\{f_n, g_n\}\} \cap \{|S_n| > D\max\{f_n, g_n\}\}\}$$

must be smaller than the probability of the single event $\{|S_n| > Dg_n\}$. Event $A \triangleq \{|T_n| + |S_n| > (C+D)\max\{f_n, g_n\}\}$ may be written $A \triangleq \{|T_n| + |S_n| - D\max\{f_n, g_n\} > C\max\{f_n, g_n\}\}$. However, if event $B \triangleq \{|S_n| \leq D\max\{f_n, g_n\}\}$ also occurs, then $|S_n| - D\max\{f_n, g_n\} \leq 0$, so replacing this term by zero can only increase the probability. i.e. $A \cap B \subseteq E$, where $E = \{|T_n| > C\max\{f_n, g_n\}\}$, and the conclusion follows because $\operatorname{pr}(A \cap B) \leq \operatorname{pr}(E)$.

2. For any $\delta > 0$, there exists a $C < \infty$ and $0 < n_1 < \infty$ such that $\operatorname{pr}(|T_n| > Cf_n) < \frac{1}{2}\delta$ for all $n \geq n_1$. For any $\varepsilon > 0$, $\delta > 0$, there exists an $n_2 > 0$ such that

$$\operatorname{pr}(|S_n| > \varepsilon g_n/C) < \delta/2 \quad \text{for all } n > n_2.$$

Write $\varepsilon f_n g_n = \left(\frac{\varepsilon}{C} g_n\right) \cdot (C f_n)$. The result follows because

$$\text{pr}(|T_n S_n| > \varepsilon f_n g_n) \leq \text{pr}(|T_n| > C f_n) + \text{pr}(|S_n| > \varepsilon g_n / C).$$

□

The probability of an event A, $\text{pr}(A)$, is the expectation under the probability measure pr of the indicator function of A, written $\mathbb{1}(A)$, taking value 1 if event A occurs and 0 otherwise. The proof strategy for the various components of proposition 1.1.8 is useful in any context involving indicator functions.

1.2 Convergent, divergent and asymptotic series

Let f be a function of a real-valued variable $x \in (a, b)$. This function might be known or unknown. A series expansion for $f(x)$ is of the form $\sum_{k=0}^{\infty} f_k(x)$. Define the partial sum $S_K(x) = \sum_{k=0}^{K} f_k(x)$ and the remainder $R_K(x) = f(x) - S_K(x)$. For a series to be useful, $S_K(x)$ must in some sense approximate $f(x)$ arbitrarily well, i.e. it must be possible to make the remainder arbitrarily small. There are different ways in which this can be assured.

Definition 1.2.1. A series $\sum_{k=0}^{\infty} f_k(x)$ is *convergent* to $f(x)$ if, for any $\varepsilon > 0$ there exists a sufficiently large K_0 such that $|f(x) - S_K(x)| < \varepsilon$ for all $K > K_0$.

Some remarks:

- In general K_0 depends on x. If $K_0 < \infty$ can be chosen independently of $x \in (a, b)$ then the series is said to converge uniformly to $f(x)$ on (a, b).
- Convergence requires that for *fixed* x, the remainder $R_K(x)$ can be made arbitrarily small by including sufficiently many terms K.

A necessary condition for convergence is that $f_k(x) \to 0$ as $k \to \infty$. In other words, if $\lim_{k\to\infty} f_k(x) \neq 0$, then the series is divergent. This is not a sufficient condition. The *ratio test* determines whether a series is divergent or convergent.

Proposition 1.2.2. Consider $x \in (a, b)$ fixed. Suppose that $M(x) \triangleq \lim_{k\to\infty} |f_{k+1}(x)/f_k(x)| < \infty$. If $M(x) < 1$, the series is convergent; if $M(x) > 1$ it is divergent.

Usually $f_k(x)$ is chosen to be proportional to a relatively simple function $\psi_k(x)$ called the *base function* or *gauge function*. For instance, $f_k(x) = a_k(x - x_0)^k$ gives rise to a *power series*. Applying the above proposition (the ratio test) to the power series $\sum_{k=0}^{\infty} a_k(x - x_0)^k$ we find that the series is convergent for any x in the interval $\{x : |x - x_0| < R\}$, where $R = \lim_{k\to\infty} |a_k/a_{k+1}|$ is the *radius of convergence* of the series. In a *Taylor series* expansion the coefficient a_k in the above power series is $a_k = f^{(k)}(x_0)/k!$, where $f^{(k)}$ is the kth derivative of the function f to be approximated. Thus $|a_k/a_{k+1}| \propto (K + 1)!/K! = K + 1$ and

$$f(x) = \sum_{k=0}^{K} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + R_K(x). \tag{1.2.1}$$

The remainder can be expressed as

$$R_K(x) = \frac{f^{(K+1)}(\bar{x})}{(K + 1)!}(x - x_0)^{K+1},$$

where $\bar{x}$ is between x and x_0. Provided that $|f^{(K+1)}(\bar{x})| < B$, we have the following upper bound for the remainder

$$|R_K(x)| \le \frac{B|x - x_0|^{K+1}}{(K + 1)!}, \tag{1.2.2}$$

which vanishes as $K \to \infty$. However, a series need not be convergent to be useful.

Consider series (1.2.1) from a different perspective. For any fixed K, R_K can be made arbitrarily small by making $|x - x_0|$ sufficiently small. That is, for any $\varepsilon > 0$, there exists a $\delta > 0$ such that $|R_K(x)| < \varepsilon$ for all x satisfying $|x - x_0| < \delta$. More specifically,

$$\delta = \left\{\frac{\varepsilon(K + 1)!}{B}\right\}^{1/(K+1)}.$$

Exercise 1 . Show that $\lim_{x\to x_0} |R_K(x)/f_K(x)| = 0$ and $\lim_{x\to x_0} |f_{K+1}(x)/f_K(x)| = 0$ for series (1.2.1).

Assuming that the last term $f_K(x)$ included in the sum $S_K(x)$ is non-zero, we can show that $\lim_{x\to x_0} |R_K(x)/f_K(x)| = 0$, so that the Taylor series expansion is an *asymptotic series* or *asymptotic expansion* for $x \to x_0$ as well as a convergent series. A series need not be convergent for it to be asymptotic.

Definition 1.2.3. The expansion $\sum_{k=0}^{K} f_k(x)$ is an *asymptotic approximation* to $f(x)$ as $x \to x_0$ if, for all $m \le K$,

(i) $\lim_{x\to x_0} |f_{m+1}(x)/f_m(x)| = 0$;

(ii) $\lim_{x\to x_0} |\{f(x) - \sum_{k=0}^{m} f_k(x)\}/f_m(x)| = 0$

If the above conditions are satisfied, we write $f(x) \simeq \sum_{k=0}^{K} f_k(x)$ $(x \to x_0)$ and the symbol $\simeq$ means that the left and right hand sides are asymptotically equivalent as $x \to x_0$. An asymptotic expansion does not, in general, uniquely determine a function.

Condition (i) says that the terms become progressively smaller in the limit $x \to x_0$ while (ii) specifies that the remainder be less than the last term included in the partial sum. If the expansion can, in principle, be carried to an arbitrary order then we have an *asymptotic series.*

Condition (i) is guaranteed if $f_k(x) = a_k\psi_k(x)$, where the sequence of base functions $\{\psi_k(x)\}_{k\in\mathbb{N}}$, form an *asymptotic sequence*, i.e., they satisfy

$$\lim_{x\to x_0} \left|\frac{\psi_{k+1}(x)}{\psi_k(x)}\right| = 0, \quad k = 0, 1, 2, \ldots.$$

While the Taylor series, with $\psi_k(x) = (x - x_0)^k$, is the simplest type of expansion, one can expand in other functions. For instance, $\psi_k(x) = e^{ikx}$ (Fourier series), or $\psi_k(x)$ might be a kth order Bessel function, a kth order orthogonal polynomial, the kth eigenfunction of a particular integral operator, etc. The choice is one of mathematical convenience in the problem at hand. For instance, if x is sample size or more generally amount of information, it is helpful to define $z = 1/x$ so that z small corresponds to x large. A standard Taylor series expansion around $z = 0$ produces an expansion in inverse powers of x. Some series may converge in regions of interest where others do not.

1.3 Construction of asymptotic expansions

By Definition 1.2.3, $f(x) = \sum_{k=0}^{K} a_k\psi_k(x) + o\{\psi_K(x)\}$. Thus, if one takes only the leading term in this expansion,

$$f(x) = a_0\psi_0(x) + o\{\psi_0(x)\}.$$

Divide both sides by ψ_0. Then

$$\frac{f(x)}{\psi_0(x)} = a_0 + \frac{o\{\psi_0(x)\}}{\psi_0(x)},$$

so that taking the limit of the right hand side as $x \to x_0$, gives an expression for the coefficient a_0 on ψ_0. Here x_0 must be one for which Definition 1.2.3 holds. This process can be repeated to calculate each successive term in the expansion. Specifically, for the jth term we have

$$a_j = \lim_{x \to x_0} \frac{f(x) - \sum_{k=0}^{j-1} a_k \psi_k(x)}{\psi_j(x)}.$$

The expansion so obtained turns out to be unique for the particular asymptotic sequence $\{\psi_k(x)\}_{k \in \mathbb{N}}$ used. When evaluating the limits, indeterminate forms are encountered, which can often be handled through an application of L'Hôpital's rule. This states that if either $h(x) \to 0$ and $g(x) \to 0$ or $h(x) \to \pm\infty$ and $g(x) \to \pm\infty$ as $x \to x_0$, then

$$\lim_{x \to x_0} \frac{h(x)}{g(x)} = \lim_{x \to x_0} \frac{h'(x)}{g'(x)}.$$

Exercise 2 . Find an asymptotic expansion of $f(x) = \tan x$ as $x \to 0$ in terms of the base functions (a) $\psi_k(x) = x^k$, and (b) $\psi_k(x) = \sin^k(x)$.

1.4 Gamma and related integrals

The gamma function is defined as an integral of elementary functions:

$$\Gamma(x) \triangleq \int_0^\infty e^{-t} t^{x-1} dt.$$

Similarly, the beta function is

$$B(x, y) \triangleq \int_0^1 t^{x-1}(1-t)^{y-1} dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

Integration by parts shows that the gamma function satisfies $\Gamma(x) = (x-1)\Gamma(x-1)$, thus if $x = k$ is a non-negative integer, $\Gamma(k) = (k-1)!$. Some specific values of the gamma function are $\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$ and $\Gamma(2) = 1$. We will later derive an explicit asymptotic approximation to the gamma function, for $x \gg 1$, in terms of elementary functions,

$$\Gamma(x) \simeq x^x e^{-x} x^{-1/2} \sqrt{2\pi}, \tag{1.4.1}$$

from which it can be shown that, for large x, $\Gamma(x+a)/\Gamma(x) \simeq x^a$. Equation (1.4.1) is *Stirling's approximation.*

Example 1.4.1. Let R be the cosine of the angle between two independent random vectors in p-dimensional space. The density function of R is (Fisher, 1915),

$$f(r) = \frac{\Gamma(p/2)}{\sqrt{\pi}\Gamma\{(p-1)/2\}}(1-r^2)^{(p-3)/2}, \quad -1 < r < 1.$$

Notice that $(1-r^2)^{(p-3)/2} \to 0$ as $p \to \infty$ for all $r \neq 0$ and is 1 at $r = 0$, i.e., at angle $\cos^{-1}(0) = \pi/2$, showing that random vectors are nearly orthogonal in high-dimensions.

Exercise 3 . Calculate the variance of R from example 1.4.1.

Exercise 4 . Let X be a random variable with density $f(x) = (1-\gamma)x^{-\gamma}$ $(0 < x < 1, 0 < \gamma < 1)$ and let $M = \min\{V_1, \ldots, V_{k-1}\}$, where the V_i $(i = 1, \ldots, k-1)$ are independent uniformly distributed random variables on $(0, 1)$. What is the probability that X is smaller than M? What is this probability in the limit as $\gamma \to 0$?

The motivation behind Question 4 will be given in the solutions.

Exercise 5 . It is clear, by symmetry, that if k is an odd positive integer,

$$\int_{-\infty}^{\infty} e^{-tx^2} x^k dx = 0, \quad t > 0,\ k = 1, 3, 5, \ldots.$$

By writing $k = 2n$, obtain an explicit representation of this integral for even positive integers.

Exercise 6 . Suppose that the frequency with which a random variable X falls within the infinitesimal range dx is

$$f_X(x)dx = \frac{\Gamma\{(n+1)/2\}}{\Gamma(n/2)\sqrt{\pi n}}(1 + x^2/n)^{-(n+1)/2}dx,$$

i.e., f_X is the density function of a student t random variable with n degrees of freedom. Obtain an expansion of the logarithm of the density in powers of n^{-1}. Use this to show that, for large n, the student t density function coincides with the standard normal density function.

1.5 Some important generating functions

Generating functions arise in statistics primarily as a device for simplifying calculations involving sums of random variables. Information about an object of interest is encapsulated within a single function that may be expanded as a power series to recover that information without loss. Any generating function of the sequence $(a_k)_{k\in\mathbb{Z}}$ is of the form

$$G(z) = \sum_{k=0}^{\infty} a_k z^k.$$

It may be differentiated j times with respect to z and evaluated at $z = 0$ to recover a_j. We provide some examples below.

If X is a discrete random variable, taking values on the non-negative integers, then the probability generating function of X is

$$G_X(\zeta) = \mathbb{E}(\zeta^X) = \sum_{x=0}^{\infty} p_X(x)\zeta^x,$$

where $p_X(x)$ is the probability mass at x. The whole probability mass function of X is recovered by taking derivatives with respect to ζ and evaluating at $\zeta = 0$.

Exercise 7 . Let V_{S1} be binomially distributed of parameter θ_{S1} and index v_{S0}. Conditionally on any $V_{S1} = v_{S1}$, let V_{S2} be binomially distributed of parameter $\theta_{S2.1}$ and index v_{S1}. What is the unconditional distribution of V_{S2}?

Exercise 8 . Let V_{N1} be Poisson distributed of mean μ_{N1}. Conditionally on any $V_{N1} = v_{N1}$, let V_{N2} be binomially distributed of parameter $\theta_{N2.1}$ and index v_{N1}. What is the unconditional distribution of V_{N2}?

The motivation behind Questions 7 and 8 will be given in the solutions.

Let $S_n = \sum_{i=1}^{n} X_i$ where X_i is a discrete random variable with probability generating function G_{X_i}, then

$$G_{S_n}(\zeta) = \mathbb{E}(\zeta^{S_n}) = \mathbb{E}(\zeta^{X_1}\cdots\zeta^{X_n}) = \prod_{i=1}^{n} G_{X_i}(\zeta).$$

or $\{G_{X_1}(\zeta)\}^n$ if the X_i are identically distributed. The probability generating function of a weighted sum or a difference can be similarly defined.

For X an arbitrary random variable, its moment generating function is $M_X(t) = \mathbb{E}(e^{tX})$. Uncentered moments of all orders are recovered by taking derivatives with respect to t and evaluating at $t = 0$, k derivatives being required for the kth moment:

$$\frac{d^k}{dt^k} M_X(t) = \int \frac{d^k}{dt^k} e^{tx} f_X(x)dx = \int x^k e^{tx} f_X(x)dx = \mathbb{E}(X^k e^{tX}),$$

so that evaluation at $t = 0$ recovers the uncentered moments $\nu_k = \mathbb{E}(X^k)$. It follows that the Taylor series expansion of M_x about $t = 0$ is

$$M_X(t) = 1 + \sum_{k=1}^{\infty} \nu_k t^k / k!$$

Similarly, the moment generating function of $X - \mathbb{E}(X)$ can be expanded in a Taylor series about $t = 0$, giving

$$M_{X-\mu}(t) = 1 + \sum_{k=1}^{\infty} \mu_k t^k / k!$$

where $\mu_k \triangleq \mathbb{E}(X - \mu)^k$ and $\mu \triangleq \nu_1$, so that $\mu_1 = 0$ and the summation can be started at $k = 2$. The cumulant generating function of X, denoted by $K_X(t)$ is defined as the logarithm of the moment generating function, with Taylor series expansion defining the *cumulants* $(\kappa_m)_{m \in \mathbb{N}}$ of X as

$$K_X(t) = \sum_{m=0}^{\infty} \kappa_m t^m / m!.$$

To deduce the relationship between the moments and the cumulants, write

$$\begin{aligned} K_X(t) &= \sum_{m=0}^{\infty} \kappa_m t^m / m! = \log M_X(t) = \log\{e^{t\mu} e^{-t\mu} M_X(t)\} \\ &= \log\{e^{t\mu} M_{X-\mu}(t)\} = t\mu + \log\Big(1 + \sum_{k=2}^{\infty} \mu_k t^k / k!\Big), \end{aligned}$$

from which we immediately observe that $\kappa_0 = 0$. Expanding $\log(1 + z)$ around $z = 0$ and equating powers of the dummy variables leads to the following expressions for the cumulants in terms of the centered moments. This requires some combinatorics for compositions of series and will not be discussed here. See e.g., McCullagh (1986 or 2017), or Comtet (1974).

$$\kappa_1 = \mu, \quad \kappa_2 = \mu_2 \triangleq \sigma^2, \quad \kappa_3 = \mu_3, \quad \kappa_4 = \mu_4 - 3\mu_2^2 \tag{1.5.1}$$

$$\kappa_5 = \mu_5 - 10\mu_3\mu_2, \quad \kappa_6 = \mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 10\mu_2^3. \tag{1.5.2}$$

The inverse relations are

$$\mu_4 = \kappa_4 + 3\kappa_2^2, \quad \mu_5 = \kappa_5 + 10\kappa_3\kappa_2 \tag{1.5.3}$$

$$\mu_6 = \kappa_6 + 15\kappa_4\kappa_2 + 10\kappa_3^2 + 15\kappa_2^3. \tag{1.5.4}$$

An alternative way of deriving these expressions is to expand e^z around $z = 0$ in the expression,

$$1 + \sum_{k=2}^{\infty} \mu_k t^k/k = M_{X-\mu}(t) = \exp\{t\mu + K_X(t)\} = \exp\Big(t\mu + \sum_{m=1}^{\infty} \kappa_m t^m/m!\Big),$$

and equate coefficients on the dummy variable.

The most important moment generating function to be able to recognize is that of the standard normal law. This is calculated as follows

$$\begin{aligned} M_X(t) &= \int_{\mathbb{R}} e^{tx} f_X(x)dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\Big\{-\frac{x^2}{2} + tx\Big\}dx \\ &= \frac{1}{\sqrt{2\pi}} \int \exp\{t^2/2\} \exp\Big\{-\frac{t^2}{2} - \frac{x^2}{2} + tx\Big\}dx \\ &= \exp\{t^2/2\} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\Big\{-\frac{1}{2}(x-t)^2\Big\}dx}_{=1} = \exp\{t^2/2\}. \end{aligned}$$

The moment generating function of a normally distributed random variable X with mean μ and variance σ^2 is simply obtained by noting that $X = \mu + \sigma Z$, where Z is standard normally distributed. Thus

$$M_X(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t} \exp\{\sigma^t t^2/2\} = \exp\{\mu t + \sigma^2 t^2/2\}.$$

Exercise 9 . Suppose that the frequency with which a random variable X falls within the infinitesimal range dx is

$$f_X(x)dx = \beta^{-1} \exp[-\{(x-\alpha)/\beta\} - e^{-(x-\alpha)/\beta}]dx.$$

Find the moment generating function of X and the first few (uncentered) moments.

Motivation for this question appears in the solution.

The moment generating function $M_X(t)$ is equivalent except for sign to a Laplace transform of the density function of X. Laplace and Fourier transforms are useful for simplifying calculations involving convolutions, because the Laplace transform of a convolution of two functions, f and

g, is the product of their Laplace transforms. Thus, remembering that the density of a sum of random variables is a convolution of their density functions, it is natural to work with the moment generating or cumulant generating function whenever sums of random variables arise.

Let $S_n = \sum_{i=1}^n X_i$, then $M_{S_n}(t) = \prod_{i=1}^n M_{X_i}(t)$, or $\{M_{X_1}(t)\}^n$ if the X_i are identically distributed. The cumulant generating function satisfies $K_{S_n}(t) = \sum_{i=1}^n K_{X_i}(t)$ or $nK_{X_1}(t)$ if the X_i are identically distributed.

Remark 1.5.1. When the integral in the expectation operator defining the moment generating function does not converge, standard practice is to work with the characteristic function $\varphi_X(t) = \mathbb{E}(e^{itX})$, $t \in \mathbb{R}$, which always exists. This has the same moment generating property as $M_X(t)$.

Other generating functions arising in statistics include the autocovariance generating functions of time-dependent processes.

ASYMPTOTIC METHODS AND STATISTICAL APPLICATIONS

Chapter 2: Asymptotic approximation of integrals

HEATHER BATTEY

FEBRUARY 21, 2020

2.1 Introduction

It is common in many mathematical fields, including statistics, to encounter integrals that do not possess an exact analytic solution.

In §2.2 and §2.3 we discuss two simple strategies for approximating general integrals of the form

$$I(\lambda) = \int_a^b g(\lambda, t)dt, \;\; I(\lambda) = \int_a^\lambda g(\lambda, t)dt, \;\; I(\lambda) = \int_\lambda^b g(\lambda, t)dt, \;\; \lambda \to \lambda_0, \tag{2.1.1}$$

where a and b are given and could be zero, finite or infinite.

Four important classes of integral, to be studied under the limiting operation $\lambda \to \infty$, are:

$$I_W(\lambda) = \int_0^b g(t)e^{-\lambda t}dt; \quad I_L(\lambda) = \int_a^b g(t)e^{\lambda u(t)}dt;$$

$$I_F(\lambda) = \int_a^b g(t)e^{i\lambda u(t)}dt; \quad I_C(\lambda) = \int_{\mathcal{C}} h(s)e^{\lambda f(s)}ds.$$

In these expressions a and b are given and could be zero, finite or infinite, and u and g are real-valued functions of a real variable t that are, by convention, independent of λ, although this is not a strict requirement. In $I_C(\alpha)$, $s = t + iy$ $(t, y \in \mathbb{R})$ is a complex variable, $f(s) = u(t, y) + iv(t, y)$, where u and v are real-valued functions of the two real-valued variables t and y, and $\mathcal{C}$ is a contour in the complex plane. These integrals will be discussed in §2.4, §2.5 and §2.6.

Equations of this type often result from using integral transforms to solve certain types of linear ordinary or partial differential equations. Similarly, calculations in statistics are often most easily performed using moment generating or characteristic functions, producing integrals of these types.

2.2 Elementary expansions of the integrand

The most elementary method for approximating integrals of the form (2.1.1) is by a Taylor series (or other series) expansion of the integrand. Some examples are:

(a) $I(\lambda) = \int_0^1 t^{-1}\sin(\lambda t)dt$, $\lambda \to 0$;

(b) $I(\lambda) = \int_\lambda^\infty e^{-t^4}dt$, $\lambda \to 0$.

(c) $I(\lambda) = \int_0^\infty e^{-\lambda t}\cos(t)dt$, $\lambda \to \infty$.

(a) Writing $x = \lambda t$, $|x| \ll 1$ for all $0 \le t \le 1$ as $\lambda \to 0$. Expanding $\sin(x)$ in a Taylor series around zero gives

$$I(\lambda) = \sum_{k=0}^{\infty}(-1)^n \frac{\lambda^{2k+1}}{(2k+1)!}\int_0^1 t^{-1}t^{2k+1}dt = \sum_{k=0}^{\infty}(-1)^k \frac{\lambda^{2k+1}}{(2k+1)!(2k+1)}. \tag{2.2.1}$$

With $f_k(\lambda) = (-1)^k\lambda^{2k+1}/\{(2k+1)!(2k+1)\}$ we have

$$M(\lambda) \triangleq \lim_{k\to\infty} |f_{k+1}(\lambda)/f_k(\lambda)| = 0$$

for any fixed λ, showing that the series expansion of (2.2.1) is convergent.

(b) Expanding e^{-t^4} directly does not work because the resulting integrals diverge. Write

$$I(\lambda) = \int_\lambda^\infty e^{-t^4}dt = \underbrace{\int_0^\infty e^{-t^4}dt}_{I_1} - \underbrace{\int_0^\lambda e^{-t^4}dt}_{I_2(\lambda)},$$

so that I_1 does not depend on λ and can be written as a gamma integral by changing variables to $v = t^4$. Thus

$$I(\lambda) = \Gamma(5/4) - I_2(\lambda).$$

Expansion of the integrand in the second term leads to

$$I_2(\lambda) = \sum_{k=0}^{\infty}\frac{1}{k!}\int_0^\lambda \frac{(-t^4)^k}{k!}dt = \sum_{k=0}^{\infty}(-1)^k\frac{\lambda^{4k+1}}{k!(4k+1)}.$$

The ratio test shows that this is a convergent series for any fixed λ.

(c) For $\lambda > 0$, the integral has the exact solution $\lambda/(1+\lambda^2)$.

Exercise 10 . Despite the availability of an exact solution, obtain an asymptotic approximation to this integral. Which factor should be expanded? Why?

2.3 Repeated integration by parts

We illustrate the technique by example.

Example 2.3.1. The upper integral $\Psi(z) = \int_z^\infty \phi(x)dx = \Phi(-z)$ of the standard normal density function $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ does not have a closed-form formula. Write

$$\begin{aligned}\Psi(z) &= \int_z^\infty \phi(x)dx = \frac{1}{\sqrt{2\pi}}\int_z^\infty \frac{-xe^{-x^2/2}}{-x}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_z^\infty -x^{-1}\left(\frac{d}{dx}e^{-x^2/2}\right)dx = \int_z^\infty \underbrace{-x^{-1}}_{u}\underbrace{d(e^{-x^2/2}/\sqrt{2\pi})}_{dv}.\end{aligned}$$

Use integration by parts: $\int udv = uv - \int vdu$, so that

$$\Psi(z) = [-\phi(x)/x]_z^\infty - \int_z^\infty \frac{\phi(x)}{x^2}dx. \tag{2.3.1}$$

The first term in (2.3.1) is $\phi(z)/z$ and the second term is of smaller order, i.e., $\Psi(z) = \phi(z)/z + o\{\phi(z)/z\}$, $(z \to \infty)$. To see this, multiply and divide the second term in (2.3.1) by x and notice that $x^{-3} \leq z^{-3}$ for all $x > z$, so that

$$\int_z^\infty \frac{\phi(x)}{x^2}dx \leq \frac{1}{z^3}\int_z^\infty x\phi(x)dx = \frac{\phi(z)}{z^3} = o\left\{\frac{\phi(z)}{z}\right\}, \quad z \to \infty.$$

Repetition of this process, with dv the same in every application and u appropriately redefined gives the improved approximation

$$\Psi(z) = \frac{\phi(z)}{z}\left\{1 - \frac{1}{z^2} + \frac{3}{z^4} - \frac{3.5}{z^6} + \cdots\right\}. \tag{2.3.2}$$

While, for fixed z, the series is divergent, for $z \to \infty$, equation (2.3.2) defines an asymptotic series with leading term $\phi(z)/z$ and base functions $\{1, 1/z^2, 1/z^4, \ldots\}$. The remainder from truncating at an arbitrary term decays z^2 times as fast $(z \to \infty)$ as the last term retained in the partial sum. The terms alternate in sign and the corresponding partial sums provide alternately upper and lower bounds on $\Psi(z)$.

Exercise 11 . The incomplete gamma function is given by

$$\gamma_m(z) = \frac{1}{\Gamma(m)} \int_z^\infty x^{m-1} e^{-x} dx.$$

Use repeated integration by parts to obtain an asymptotic series expansion for $\gamma_m(z)$ as $z \to \infty$.

2.4 Watson's lemma and Laplace's method

2.4.1 Watson's lemma

Watson's lemma (Watson, 1944, p236) provides a general asymptotic expansion for integrals of the form $I_W(\lambda) = \int_0^b g(t)e^{-\lambda t}dt$. Such integrals are similar to those of type $I_L(\lambda) = \int_a^b g(t)e^{\lambda u(t)}dt$, to be discussed in §2.4. The formal justification for the methods discussed there is provided by arguments analogous to those used in the proof of Watson's lemma.

Proposition 2.4.1 (Watson's lemma)**.** Let $0 < b \leq \infty$. Suppose that the function $g(t)$ has an asymptotic expansion of the form

$$g(t) \simeq \sum_{k=0}^{K} \alpha_k t^{\beta_k}, \quad t \to 0$$

with $-1 < \beta_0 < \beta_1 < \cdots < \beta_K$. Then the integral $I_W(\lambda) = \int_0^b g(t)e^{-\lambda t}dt$ has an asymptotic expansion

$$\begin{aligned} I_W(\lambda) &= \int_0^\infty e^{-\lambda t} \sum_{k=0}^{K} \alpha_k t^{\beta_k} dt + O\{\lambda^{-(\beta_K+1)}\} \\ &= \sum_{k=0}^{K} \alpha_k \lambda^{-(\beta_k+1)} \Gamma(\beta_k + 1) + O\{\lambda^{-(\beta_K+1)}\}, \quad \lambda \to \infty. \end{aligned}$$

Proof. Write $\widetilde{I}_W(\lambda) \triangleq \sum_{k=0}^{K} \alpha_k \lambda^{-(\beta_k+1)} \Gamma(\beta_k + 1)$. Then

$$\begin{aligned} \left| I_W(\lambda) - \widetilde{I}_W(\lambda) \right| &= \left| I_W(\lambda) - \sum_{k=0}^{K} \alpha_k \int_0^\infty t^{\beta_k} e^{-\lambda t} dt \right| \\ &\leq \underbrace{\left| \int_0^\varepsilon g(t)e^{-\lambda t}dt - \int_0^\infty e^{-\lambda t} \sum_{k=0}^{K} \alpha_k t^{\beta_k} dt \right|}_{\mathrm{I}_1} + \underbrace{\left| \int_\varepsilon^b g(t)e^{-\lambda t}dt \right|}_{\mathrm{I}_2}. \end{aligned}$$

We have $\mathrm{I}_2 \leq Ge^{-\lambda\varepsilon}$ by Hölder's inequality, where $G = \int_\varepsilon^b |g(t)|dt$.

$$\mathrm{I}_1 \leq \underbrace{\left|\sum_{k=0}^{K} \alpha_k \left(\int_0^\varepsilon - \int_0^\infty\right) t^{\beta_k} e^{-\lambda t} dt\right|}_{\mathrm{I}_{11}} + \underbrace{\left|\int_0^\varepsilon R(t) e^{-\lambda t} dt\right|}_{\mathrm{I}_{12}},$$

where $R(t) = g(t) - \sum_{k=0}^{K} \alpha_k t^{\beta_k}$ and is, by the definition of an asymptotic expansion, of smaller order than the last term in the series. Thus $R(t) = o(t^{\beta_K})$ or, equivalently, there exists a $0 < C < \infty$ and $0 < \delta < b$ such that $|R(t)| < Ct^{\beta_K}$ for all $0 < t < \delta$. It follows that

$$\mathrm{I}_{12} \leq C \int_0^\varepsilon t^{\beta_K} e^{-\lambda t} dt \leq C \int_0^\infty t^{\beta_K} e^{-\lambda t} dt = C\lambda^{-(\beta_K+1)}\Gamma(\beta_K + 1).$$

This tends to zero as $\lambda \to \infty$ because $\beta_K > -1$ by the statement of the proposition.

For the final term, I_{11},

$$\mathrm{I}_{11} = \left|\sum_{k=0}^{K} \alpha_k \int_\varepsilon^\infty t^{\beta_k} e^{-\lambda t} dt\right| \leq \sum_{k=0}^{K} |\alpha_k| \int_\varepsilon^\infty t^{\beta_k} e^{-\lambda t} dt.$$

The same strategy as for I_{12} could be used, resulting in a sum of terms that depends on λ. But a better approach, which makes the rate of convergence in proposition 2.4.1 explicit, is to write $e^{-\lambda t} = e^{-qt} e^{-(\lambda-q)t}$. Thus $e^{-\lambda t} \leq e^{-qt} e^{-(\lambda-q)\varepsilon}$ provided that $\lambda > q$ and $t \geq \varepsilon$. By assumption $\lambda \gg 1$, so take $q = 1$ leading to

$$\mathrm{I}_{11} \leq e^{-(\lambda-1)\varepsilon} \sum_{k=0}^{K} |\alpha_k| \int_0^\infty e^{-t} t^{\beta_k} dt.$$

Both I_2 and I_{11} tend to zero exponentially fast as $\lambda \to \infty$. provided that ε is chosen not to be of smaller order than $1/\lambda$. The dominant order term as $\lambda \to \infty$ is I_{12}, so that

$$\left|I_W(\lambda) - \widetilde{I}_W(\lambda)\right| = O\{\lambda^{-(\beta_K+1)}\}, \quad \lambda \to \infty.$$

□

The following version of Watson's lemma can be similarly proved.

Proposition 2.4.2. Suppose that $g(t)$ has an asymptotic expansion at zero of the form

$$g(t) \simeq \sum_{k=0}^{K} \frac{g^{(k)}(0)}{k!} t^k, \quad t \to 0.$$

Let $\alpha_k \triangleq g^{(k)}(0)/k!$. Then

$$I_W(\lambda) = \int_0^b t^\gamma g(t) e^{-\lambda t^\nu} dt = \sum_{k=0}^{K} \frac{\alpha_k}{\nu} \Gamma\left(\frac{\gamma+k+1}{\nu}\right) \lambda^{-(\gamma+k+1)/\nu} + O\left\{\lambda^{-(\gamma+K+2)/\nu}\right\}, \quad \lambda \to \infty.$$

Exercise 12 . Consider the integral

$$\int_0^\infty \frac{e^{-\lambda \cosh t}}{\sin^{1/3} t} dt, \quad \lambda \to \infty. \tag{2.4.1}$$

By a suitable change of variable, write (2.4.1) in a form to which proposition 2.4.1 or 2.4.2 can be applied.

2.4.2 Laplace's method

The key idea of the previous result is that, since the integral is dominated by the contribution in a neighbourhood of $t = 0$, we can replace the function g by its asymptotic expansion at $t = 0$. The original range of integration $(0, b)$ can be extended to the range $(0, \infty) = (0, \varepsilon) \cup (\varepsilon, \infty)$ because the contribution from the (ε, ∞) is small. On the larger range of integration the resulting integral is analytically tractable.

Watson's lemma deals with integrals in which a neighbourhood of zero dominates the integral. In fact, many integrals in which an endpoint dominates can be expressed in the form $I_W(\lambda)$ by a change of variables.

Consider now integrals of the form $I_L(\lambda) = \int_a^b g(t) e^{\lambda u(t)} dt$, where $u(t)$ achieves a maximum at a point c, say, where $a < c < b$, so that a neighbourhood of c dominates $I_L(\lambda)$ for large λ, provided that $g(t)$ is not zero near c.

Exercise 13 . Let $X^{(1)} = \max\{X_1, \ldots, X_n\}$ and $X^{(n)} = \min\{X_1, \ldots, X_n\}$, the largest and smallest order statistics in a sample of size n, and let the X_i be independent and identically distributed with distribution function F and density function f. It can be shown that the frequency with which $(X^{(1)}, X^{(n)})$ falls in the infinitesimal area $dx_{(1)} dx_{(2)}$ is

$$n(n-1)[F\{x_{(1)}\} - F\{x_{(n)}\}]^{n-2} f\{x_{(1)}\} f\{x_{(n)}\} dx_{(1)} dx_{(n)}. \tag{2.4.2}$$

The density function of $R_n = X^{(1)} - X^{(n)}$ can be obtained as follows. Write $X^{(n)} = U - \frac{1}{2} R_n$ and $X^{(1)} = U + \frac{1}{2} R_n$, where $U = \{X^{(1)} + X^{(n)}\}/2$; change variables $(X^{(1)}, X^{(n)}) \to (U, R_n)$, with absolute Jacobian determinant 1; integrate over all permissible values of U. Thus we have

$$f_{R_n}(r) = n(n-1)\int_{-\infty}^{\infty}\{F(u+\tfrac{1}{2}r) - F(u-\tfrac{1}{2}r)\}^{n-2} f(u+\tfrac{1}{2}r) f(u-\tfrac{1}{2}r)du. \tag{2.4.3}$$

Show that the right hand side of (2.4.3) can be written in the form $I_L(\lambda)$.

We write

$$I_L(\lambda) = \int_a^{c-\varepsilon} + \int_{c-\varepsilon}^{c+\varepsilon} + \int_{c+\varepsilon}^{b} \triangleq I_L^-(\lambda;\varepsilon) + I_L(\lambda;\varepsilon) + I_L^+(\lambda;\varepsilon),$$

where $0 < \varepsilon \ll 1$. It can be shown by analogous arguments to those appearing in the proof of Watson's lemma that I_L^- and I_L^+ are exponentially small compared with $I_L(\lambda;\varepsilon)$ and therefore that $I_L(\lambda) \simeq I_L(\lambda;\varepsilon)$. To evaluate $I_L(\lambda;\varepsilon)$, we approximate g and u locally near c by Taylor series expansion. This technique of approximating integrals of the form $I_L(\lambda)$ is called *Laplace's method.*

$$u(t) = u(c) + u'(c)(t-c) + \tfrac{1}{2}u''(c)(t-c)^2 + O\{(t-c)^3\}.$$

Since c is a point of maximum of u, $u'(c) = 0$ and $u''(c) < 0$. We also have $g(t) = g(c) + O(t-c)$. Thus

$$I_L(\lambda;\varepsilon) \simeq g(c)e^{\lambda u(c)} \int_{c-\varepsilon}^{c+\varepsilon} \exp\left\{\lambda u''(c)(t-c)^2/2\right\} dt.$$

It is natural to change variables to $s = \{-\lambda u''(c)\}^{1/2}(t-c)$. Thus, recalling that $u''(c) < 0$, $dt = \{-\lambda u''(c)\}^{-1/2}ds$, from which we obtain

$$I_L(\lambda;\varepsilon) \simeq \frac{g(c)e^{\lambda u(c)}}{\{-\lambda u''(c)\}^{1/2}} \int_{-\{-\lambda u''(c)\}^{1/2}\varepsilon}^{\{-\lambda u''(c)\}^{1/2}\varepsilon} e^{-s^2/2} ds.$$

Provided that $\lambda^{-1/2} \ll \varepsilon \ll 1$, the limits of integration tend to $(-\infty, \infty)$ as $\lambda \to \infty$, thus the leading term approximation to $I_L(\lambda;\varepsilon)$ is

$$I_L(\lambda;\varepsilon) \simeq \left\{\frac{2\pi}{-\lambda u''(c)}\right\}^{1/2} g(c)e^{\lambda u(c)}, \quad (\lambda \to \infty).$$

Finding higher order terms is more difficult and less elegant. However, often the first order term is sufficiently accurate, even for relatively small values of λ. Numerical work will check the quality of the approximation.

Exercise 14 . Use Laplace's method derive Stirling's approximation to the gamma function, i.e., $\Gamma(x) \simeq x^x e^{-x} x^{-1/2} \sqrt{2\pi}$.

Exercise 15 . Use Laplace's method to obtain an approximation to

$$I(\lambda) = \int_0^\pi e^{\lambda \sin t} dt, \quad (\lambda \to \infty).$$

If the maximum of u is attained at an end point rather than in the interior of (a, b) the key difference is that typically $u'(a) < 0$ and $u'(b) > 0$ if the maximum is achieve at b.

Exercise 16 . Construct a general approximation to an integral of the form

$$I_L(\lambda) = \int_a^b g(t) e^{\lambda u(t)} dt$$

when the maximum of u is attained at a with (i) $u'(a) < 0$; (ii) $u'(a) = 0$ and $u''(a) < 0$.

2.5 Method of stationary phase

This section deals with so called *generalized Fourier integrals* of the form

$$I_F(\lambda) = \int_a^b g(t) e^{i\lambda u(t)} dt,$$

and their approximation for large λ. Integrals of this type arise upon Fourier transformation and are therefore encountered when solving certain types of differential equations. In statistics they arise when calculating characteristic functions. They are related to the integrals of type $I_C(\lambda)$ to be discussed in the next section.

In $I_F(\lambda)$, $u(t)$ is a real-valued function, called the *phase function*, of the real variable t, thus $e^{i\lambda u(t)}$ is constrained to the unit circle in the complex plane and its real and imaginary parts change sign rapidly as a function of t for large λ.

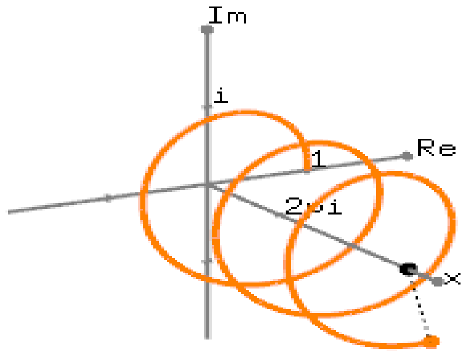

Figure 1: Graph of e^{ix}.

The last statement can be seen by decomposing $e^{i\lambda u(t)}$ into its real and imaginary parts using Euler's representation, specifically

$$e^{i\lambda u(t)} = \cos\{\lambda u(t)\} + i\sin\{\lambda u(t)\}.$$

The rapid oscillations of the intregrand produce almost complete cancellation upon integration, so that the main contribution to integrals of the form $I_F(\lambda)$ come from the end points a and b, and a neighbourhood of the points at which the phase function $u(t)$ is stationary, i.e., has zero derivative.

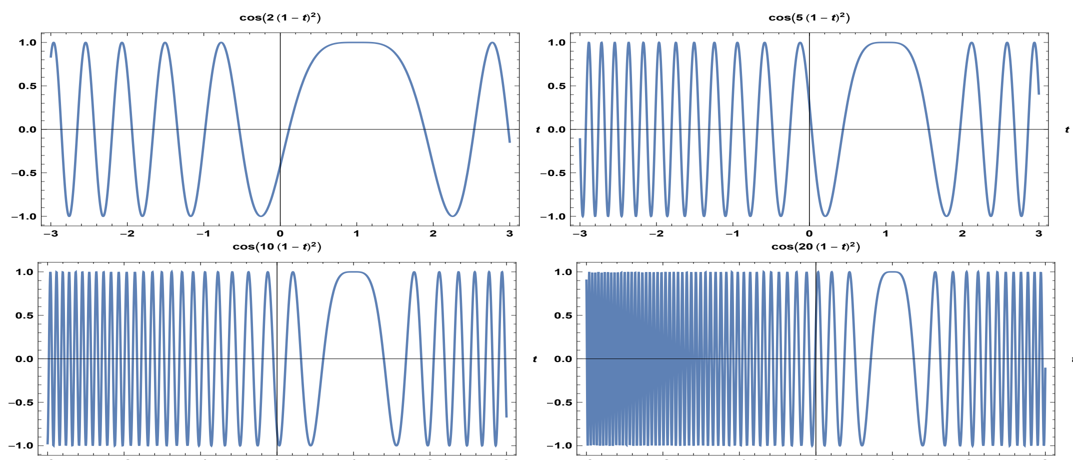

Figure 2: The function $\cos\{\lambda(1-t)^2\}$ for $\lambda = 2, 5, 10, 20$.

As in Laplace's method, decompose the integral as

$$I_F(\lambda) = \int_a^{c-\varepsilon} + \int_{c-\varepsilon}^{c+\varepsilon} + \int_{c+\varepsilon}^{b} \triangleq I_F^- + I_F(\lambda;\varepsilon) + I_F^+,$$

Suppose that u has a single stationary point at $c \in (a,b) \subset \mathbb{R}$. Then, for some $\varepsilon \ll 1$, assuming

$u''(c) \neq 0$,

$$\begin{aligned} I_F(\lambda) &\simeq I_F(\lambda;\varepsilon) \simeq \int_{c-\varepsilon}^{c+\varepsilon} g(c)\exp[i\lambda\{u(c)+\tfrac{1}{2}u''(c)(t-c)^2\}]dt \\ &= e^{i\lambda u(c)}g(c)\int_{c-\varepsilon}^{c+\varepsilon}\exp\{\tfrac{i\lambda}{2}u''(c)(t-c)^2\}dt. \end{aligned}$$

Change variables to $s = (t-c)\{\frac{1}{2}\lambda|u''(c)|\}^{1/2}$ so that

$$\begin{aligned} I_F(\lambda) &\simeq e^{i\lambda u(c)}g(c)\left\{\frac{2}{\lambda|u''(c)|}\right\}^{1/2}\int_{-\{\frac{1}{2}\lambda|u''(c)|\}^{1/2}\varepsilon}^{\{\frac{1}{2}\lambda|u''(c)|\}^{1/2}\varepsilon}\exp[i\mathrm{sgn}\{u''(c)\}s^2]ds \\ &\simeq 2e^{i\lambda u(c)}g(c)\left\{\frac{2}{\lambda|u''(c)|}\right\}^{1/2}\underbrace{\int_0^\infty \exp[i\mathrm{sgn}\{u''(c)\}s^2]ds}_{J}. \end{aligned}$$

The function $\exp\{\pm iu''(c)s^2\}$ is analytic. Since integrals of analytic functions are path independent, we will treat s as complex and use Cauchy's theorem of complex analysis to evaluate J. To this end, let $\gamma(R) = C_1(R)\cup C_2(R)\cup C_3(R)$ be a closed curve in the complex plane, where $C_1(R)$ is from the origin to a point R along the real axis, $C_2(R)$ is from R to $Re^{i\theta}$ for an angle θ to be chosen in due course, and $C_3(R)$ is from $Re^{i\theta}$ back to the origin.

The sign function in J is dealt with by considering positive and negative cases separately. For the positive case, J^+ say,

$$J^+ = \lim_{R\to\infty}\int_{C_1(R)} e^{is^2}ds = \lim_{R\to\infty}\left(\oint_{\gamma(R)} e^{is^2}ds - \int_{C_2(R)} e^{is^2}ds - \int_{C_3(R)} e^{is^2}ds\right).$$

By Cauchy's theorem, the first term in parenthesis is zero for any R. Parameterize the second path as $C_2(R) = \{s = Re^{it} : t\in[0,\theta]\}$ so that, on the path, $s^2 = R^2e^{2it}$, $t\in[0,\theta]$. The modulus of the second integral is

$$\begin{aligned} \left|\int_{C_2(R)} e^{is^2}ds\right| &= \left|\int_0^\theta e^{iR^2\cos(2t)}e^{i^2R^2\sin(2t)}iRe^{it}dt\right| \\ &\leq R\int_0^\theta e^{i^2R^2\sin(2t)}dt \leq R\int_0^\theta e^{-R^2 4t/\pi}dt, \quad \theta\in[0,\pi/4], \end{aligned} \tag{2.5.1}$$

where the last inequality follows because the graph of $\sin(x)$ is concave on the interval $x\in[0,\pi/2]$ and so lies above the straight line connecting its end points, implying that $\sin(x)\geq 2x/\pi$ provided that $x\in[0,\pi/2]$. We therefore require that $\theta\in[0,\pi/4]$. The upper bound can be calculated explicitly as

$$\int_0^\theta e^{-R^2 4t/\pi}dt = -\left.\frac{e^{-R^2 4t/\pi}}{4R^2/\pi}\right|_{t=0}^{\theta} = \frac{\pi(1-e^{-R^2 4\theta/\pi})}{4R^2}.$$

The right hand side of (2.5.1) is $o(R^{-1})$ $(R \to \infty)$ provided that $0 \le \theta \le \pi/4$ and it follows that

$$J^+ = -\lim_{R\to\infty} \int_{C_3(R)} e^{is^2} ds = -e^{i\theta} \lim_{R\to\infty} \int_R^0 \exp[ir^2\{\cos(2\theta) + i\sin(2\theta)\}]dr.$$

We are free to choose any convenient $0 \le \theta \le \pi/4$, and a wise choice is $\theta = \pi/4$ so that the above integral is of a familiar form, i.e.,

$$J^+ = -e^{i\theta} \lim_{R\to\infty} \int_R^0 e^{-r^2 \sin(2\theta)} e^{ir^2 \cos(2\theta)} dr = e^{i\pi/4} \int_0^\infty e^{-r^2} dr = e^{i\pi/4} \frac{\sqrt{\pi}}{2}.$$

An analogous argument, taking $\theta = -\pi/4$, shows that $J^- = e^{-i\pi/4}\frac{\sqrt{\pi}}{2}$. We conclude that

$$I_F(\lambda) \simeq 2e^{i\lambda u(c)} g(c) \left\{ \frac{2}{\lambda |u''(c)|} \right\}^{1/2} e^{\mathrm{sgn}\{u''(c)\} i\pi/4} \frac{\sqrt{\pi}}{2}.$$

For the case in which

$$u'(c) = u''(c) = \cdots = u^{(p-1)}(c) = 0, \quad u^{(p)}(c) \ne 0, \quad c \in (a, b)$$

we have the following generalization of this.

$$I_F(\lambda) \simeq \begin{cases} \frac{2g(c)e^{i\lambda u(c)}}{\lambda^{1/p}} \left\{ \frac{p!}{|u^{(p)}(c)|} \right\}^{1/p} \frac{\Gamma(1/p)}{p} \exp\left[\frac{\mathrm{sgn}\{u^{(p)}(c)\} i\pi}{2p} \right], & (p \text{ even}) \\ \frac{2g(c)e^{i\lambda u(c)}}{\lambda^{1/p}} \left\{ \frac{p!}{|u^{(p)}(c)|} \right\}^{1/p} \frac{\Gamma(1/p)}{p} \cos\left[\frac{\mathrm{sgn}\{u^{(p)}(c)\} \pi}{2p} \right], & (p \text{ odd}). \end{cases}$$

For further discussion of the method of stationary phase, including the situation in which $u'(t) \ne 0$ for all $t \in (a, b)$, see Bender and Orzag (1999, §6.5).

2.6 Method of steepest descents

Both of the previous sections considered integrals over the real line. Contour integrals in the complex plane arise naturally in statistics when inverting characteristic functions to obtain probability density functions, although the method of steepest descents can also be used to approximate integrals over the real line. Characteristic functions are typically introduced for simplifying calculations involving sums of random variables. In the present section we consider asymptotic approximation of integrals of the form $I_C(\lambda) = \int_{\mathcal{C}} h(s) e^{\lambda f(s)} ds$ for $\lambda \to \infty$, where $\mathcal{C}$ is a path in the complex plane. The complex valued function f of the complex variable s is assumed to be analytic and has a real and imaginary part and it is helpful to write $f(s) = u(t, y) + iv(t, y)$, where u is called the *altitude function* and controls the size of the integrand while v is called the *phase function* and dictates the oscillation of the integrand. Here, $s = t + iy$ and u and v are real-valued functions of the complex variable s, and hence of the two real variables t and y. There are two special cases. If $s = t$, $f(s) = u(t)$, $h(s) = g(t)$, then $I_C(\lambda)$ is an integral of Laplace type, studied in §2.4. If $s = t$, $f(s) = iv(t)$, $h(s) = g(t)$, then $I_C(\lambda)$ is an integral of generalized Fourier form, discussed in §2.5.

Think of the functions $\{u(t, y) : (t, y) \in \mathcal{S} \subseteq \mathbb{R}^2\}$ and $\{v(t, y) : (t, y) \in \mathcal{S} \subseteq \mathbb{R}^2\}$ as mountains with contour lines showing the points in $\mathbb{R}^2$ of equal height above sea level. Contrary to real geography, two mountains can exist in the same place.

The altitude and the phase of f, dictated by u and v change as we move along the directed contour $\mathcal{C}$. The strategy of the method of steepest descents is to avoid the oscillations, which make the integral difficult to approximate, by deforming the integration path to a new path $\mathcal{C}'$ in a way that is permitted by Cauchy's theorem of complex analysis. In particular $\gamma = C \cup -C'$ should be a closed curve enclosing a region of analyticity of f, where the minus sign indicates a reversal of the direction of integration along C'. Because the phase is constant along $\mathcal{C}'$, the integral is

$$\int_{\mathcal{C}} h(s)e^{\lambda f(s)}ds = \int_{\mathcal{C}'} h(s)e^{\lambda f(s)}ds = e^{-i\lambda v(s_0)} \underbrace{\int_{\mathcal{C}'} h(s)e^{\lambda u(s)}ds}_{I(\lambda)} \tag{2.6.1}$$

where s_0 is an arbitrary point along C'. The integral $I(\lambda)$ is essentially of Laplace type, the difference being that integration is over a curve, which will in general need to be parameterized.

A constant phase path has some important geometric properties. Consider the families of curves, defined for any η and κ:

$$\begin{aligned} \mathcal{C}_v^\eta &= \{(t, y) : v(t, y) = \eta\}, \\ \mathcal{C}_u^\kappa &= \{(t, y) : u(t, y) = \kappa\}. \end{aligned}$$

These could be empty sets for some values of η and κ. The space $\mathbb{R}^2$ in which (t, y) resides is covered with such curves.

Consider any two such curves $\mathcal{C}_v^\eta$ and $\mathcal{C}_u^\kappa$. At every point along $\mathcal{C}_v^\eta$ there is a normal direction given by $\nabla v/\|\nabla v\|$ and similarly at every point along $\mathcal{C}_u^\kappa$. We will demonstrate that $\mathcal{C}_v^\eta$ and $\mathcal{C}_u^\kappa$ intersect at right angles for any η and κ.

Let v_t, v_y, u_t and u_y denote the partial derivatives of v and u with respect to t and y, so that

$\nabla u = (u_t, u_y)^T$ and $\nabla v = (v_t, v_y)^T$. The inner product of ∇u and ∇v at a point, s_0 say, is

$$\begin{aligned} \nabla v(s_0) \cdot \nabla u(s_0) &= v_t(s_0)u_t(s_0) + v_y(s_0)u_y(s_0) \\ &= v_t(s_0)v_y(s_0) + v_y(s_0)(-v_t)(s_0) = 0, \end{aligned} \tag{2.6.2}$$

where the penultimate equality follows from the Cauchy-Riemann equations of complex analysis: $u_t = v_y$ and $u_y = -v_t$. Equation (2.6.2) implies that ∇v and ∇u are orthogonal at every point of intersection s_0 of $\mathcal{C}_v^\eta$ and $\mathcal{C}_u^\kappa$, which in turn implies that $\nabla u(s_0)$ is tangential to $\mathcal{C}_v^\eta$ at s_0. The gradient of a function of two or more variables is the direction along which the function changes most rapidly. Thus if a mountaineer walks along a contour $\mathcal{C}_v^\eta$ of constant v, he is walking in a direction of steepest gradient of u. Since

$$\nabla(e^{\lambda u}) = \lambda e^{\lambda u} \nabla u,$$

the integrand changes most rapidly in the direction of ∇u, i.e., in the direction of the constant phase path.

By analogy with Laplace's method, we are interested in a point along the chosen constant phase path at which the integrand $I(\lambda)$ from (2.6.1) is maximized and decays sharply away from this point for large λ. As in Laplace's method, there are two cases: (i) the maximum of u along the constant phase path is at an end point of the path; (ii) the maximum is attained at a point s^* in the interior of the constant phase path.

Finding a maximum along a curve entails a directional derivative. Specifically, s^* solves

$$\nabla u \cdot \left(\frac{\nabla u}{\|\nabla u\|} \right) = 0. \tag{2.6.3}$$

It follows that, at s^*, $\|\nabla u\|^2 = 0$, i.e., $u_t = u_y = 0$, and by the Cauchy-Riemann equations $v_t = v_y = 0$ as well. The saddlepoint defining equation is

$$f'(s^*) = 0. \tag{2.6.4}$$

The formal verification of s^* as a saddlepoint of u and v is by checking the sign of the determinant of the Hessian matrix. This is, by the Cauchy-Riemann equations,

$$\begin{aligned} u_{tt}u_{yy} - u_{ty}^2 = -u_{tt}^2 - u_{ty}^2 &< 0 \\ v_{tt}v_{yy} - v_{ty}^2 = -v_{tt}^2 - v_{ty}^2 &< 0, \end{aligned}$$

where subscripts denote partial differentiation.

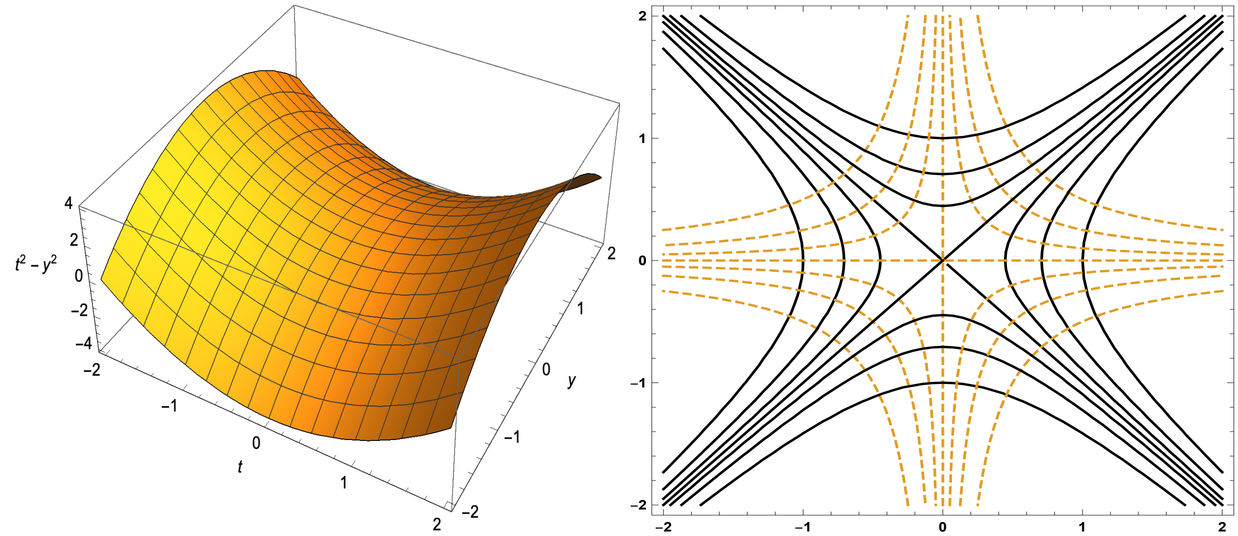

Figure 3: (left) plot of the real part $u(t,y) = t^2 - y^2$ of $f(s) = s^2 = u(t,y) + iv(t,y)$ and (right) corresponding contours of equal $u(t,y)$ (solid) and equal $v(t,y) = 2ty$ (dashed).

The local landscape around any saddlepoint is qualitatively the same as that depicted in Figure 3. In particular, there are two constant phase paths through a saddlepoint. One is a steepest ascent path from s^* and one is a steepest descent path. Care must be taken to discard the ascent path. A Laplace approximation around s^* on the steepest descent path is then performed, which is the *method of steepest descents*. We will see a key statistical example of the method of steepest descents in Chapter 3. Example 2.6.1 below is a reasonably simple example to illustrate key ideas.

As usual in complex analysis, the main difficulty in applying the method of steepest descents is in the establishment of a path that produces an easily solved integral and gives a good approximation. We have provided guidelines that lead to good approximations, but these will not always completely determine the path. It is, to some extent, a matter of imagination and experience to determine an effective path. The above discussion is oversimplified as it suggests that $\mathcal{C}$ can be deformed straightforwardly to a constant phase path $\mathcal{C}'$ that passes through a saddlepoint. This is rarely the case because the endpoints of $\mathcal{C}$ typically do not reside on the same constant phase paths. The following strategies are helpful for guiding the construction of the deformed integration contour. Let A and B denote the end points of $\mathcal{C}$.

- In the unlikely situation in which A and B reside on the same constant phase contour $\mathcal{C}_v^\eta$, $\mathcal{C}'$ can be chosen as $\mathcal{C}_v^\eta$. The value of the altitude function u at all potentially important points along $\mathcal{C}_v^\eta$ should be checked. The dominant contribution will come from the largest of these, which may be at A, B or a point in the interior of $\mathcal{C}_v^{\eta}$. Laplace's method/Watson's lemma should be applied around that point.

- If A and B live on different constant phase paths, these will never meet, and should be joined to produce a closed curve somewhere where the contribution of the integrand is small. Such joining paths are called *bypass paths.* Since constant phase paths are steepest ascent/descent paths, there will be at least one direction in which the value of u will decay sharply from the points of maximum u along the constant phase paths emanating from A and B. The integrand can often be made negligible by following such paths arbitrarily far out and constructing a simple bypath pass there. Example 2.6.1 is of this form.

- As above, A and B live on different phase paths, but there is no simple bypass path along which the integral is negligible. Identify the saddlepoint(s). There are two constant phase paths emanating from a saddlepoint. Discard the steepest ascent path and follow the steepest descent path far enough out that it can (with luck!) be joined to the constant phase paths emanating from A and B with bypass paths along which the integral is negligible.

Example 2.6.1. Consider $I_{\mathcal{C}}(\lambda) = \int_{\mathcal{C}} e^{i\lambda s^2} ds$ $(\lambda \to \infty)$ for s complex and $\mathcal{C} = [0, 1]$. This is essentially the situation considered in Figure 3 but with the roles of u and v reversed. Specifically $f(s) = is^2$ so that $u(t, y) = -2ty$ and $v(t, y) = t^2 - y^2$.

Specification of the path $\mathcal{C}'$. Solving the saddlepoint-defining equation $f'(s^*) = 0$, we obtain $s^* = 0$. The constant-v paths through s^* are thus

$$(t, y): \; v(t, y) = t^2 - y^2 = v(0, 0) = 0.$$

These are given by $\mathcal{C}_0^{(1)} : t = y$ and $\mathcal{C}_0^{(2)} : t = -y$. One of these is a steepest descent path of u and the other is a steepest ascent path. We must determine which is which by considering how u varies as we move along the paths. Along $\mathcal{C}_0^{(1)}$, $u(t, y) = -2t^2 = -2y^2$ so that $u(t, y)$ decreases as t or y move away from s^*. By contrast, along $\mathcal{C}_0^{(2)}$, $u(t, y) = 2t^2 = 2y^2$ so that the converse is true. The steepest descent path through s^* is $\mathcal{C}_0^{(1)}$. However, $\mathcal{C}_0^{(1)}$ is not a deformation of $\mathcal{C}$ as it can only share the zero endpoint. In other words, we need to specify the missing side $\mathcal{C}_1$ of a closed curve $\gamma = \mathcal{C} + \mathcal{C}_0 + \mathcal{C}_1$ so that we may take $\mathcal{C}' = \mathcal{C}_0 + \mathcal{C}_1$, where $\mathcal{C}_0$ is a suitable segment of $\mathcal{C}_0^{(1)}$.

Since v should be ideally be constant on $\mathcal{C}'$, we specify $\mathcal{C}_1$ by considering the constant-v path emanating from the upper endpoint of the integral. On such a path: $t^2 - y^2 = v(1,0) = 1$, i.e.,

$$\begin{aligned} \mathcal{C}_1^{(1)} &= \{(t,y) : t = (y^2+1)^{1/2}\} \quad \text{or} \\ \mathcal{C}_1^{(2)} &= \{(t,y) : t = -(y^2+1)^{1/2}\}. \end{aligned}$$

In order to meet $\mathcal{C}_0$, the path must be a steepest descent path from the upper endpoint, thus we take $\mathcal{C}_1$ as a suitable segment of $\mathcal{C}_1^{(1)}$. It only remains to connect the paths $\mathcal{C}_0^{(1)}$ and $\mathcal{C}_1^{(1)}$, either by finding a point at which they meet, if such a point exists, or by specifying a simple joining path:

$$\mathcal{C}^{(b)} \triangleq \{(t,b) : b \leq t \leq \sqrt{b^2+1}\}$$

with b any convenient value.

Up to the constant b, the new path has been specified as $\mathcal{C}' = \mathcal{C}_1^{(b)} \cup \mathcal{C}^{(b)} \cup \mathcal{C}_0^{(b)}$, where $\mathcal{C}_1^{(b)}$ runs from 1 to $(\{b^2+1\}^{1/2}, b)$ along $\mathcal{C}_1$, $\mathcal{C}^{(b)}$ runs horizontally in the anticlockwise direction from $(\{b^2+1\}^{1/2}, b)$ until it meets $\mathcal{C}_0$ at (b,b), and $\mathcal{C}_0^{(b)}$ runs from this point to zero along $\mathcal{C}_0$.

Parameterization of and integration along the new path. On $\mathcal{C}_0$, $t = y$ so that $s = t + it = (1+i)t$. Thus $ds = (1+i)dt$. We have

$$\begin{aligned} \int_{\mathcal{C}_0^{(b)}} e^{i\lambda s^2} ds &= (1+i)\int_b^0 e^{i\lambda(1+i)^2 t^2} dt = -(1+i)\int_0^b e^{-2\lambda t^2} dt \\ &= -(1+i)\frac{2}{\sqrt{\lambda}}\int_0^{b\sqrt{4\lambda}} e^{-x^2/2} dt \simeq -\frac{(1+i)}{2\sqrt{\lambda}}\sqrt{\frac{\pi}{2}}, \quad (\lambda \to \infty). \end{aligned}$$

On $\mathcal{C}^{(b)}$, $s = t + ib$, where $b \leq t \leq \sqrt{b^2+1}$. Therefore

$$\int_{\mathcal{C}^{(b)}} e^{i\lambda s^2} ds = \int_{\sqrt{b^2+1}}^{b} e^{i\lambda(t^2 - b^2 + 2itb)} dt = -e^{-i\lambda b^2}\int_b^{\sqrt{b^2+1}} e^{i\lambda t^2} e^{-2\lambda tb} dt$$

This can be made negligible by taking b arbitrarily large.

Finally, on $\mathcal{C}_1^{(b)}$, $t = (y^2+1)^{1/2}$, so that $s^2 = t^2 - y^2 + 2ity = 1 + iq$, where $q = 2ty = 2y\sqrt{y^2+1}$, where y runs from 0 to b. Thus

$$\int_{\mathcal{C}_1^{(b)}} e^{i\lambda s^2} ds = \frac{i}{2} e^{i\lambda}\int_0^{2b\sqrt{b^2+1}} \frac{1}{\sqrt{1+iq}} e^{-\lambda q} dq.$$

The integral is of the form $I_W(\lambda)$ discussed in §2.4 and may be evaluated using Watson's lemma. For this we require the behaviour of $(1+iq)^{-1/2}$ near $q = 0$, e.g., by Taylor series or binomial series approximation.

For an arbitrary $\alpha \in \mathbb{C}$ and $z \in \mathbb{C}$ with $|z| < 1$ the binomial series is

$$(1+z)^\alpha = \sum_{k=0}^{\infty} \frac{\alpha(\alpha-1)(\alpha-2)(\alpha-k+1)}{k!} z^k \triangleq \sum_{k=0}^{\infty} \binom{\alpha}{k} z^k$$

and

$$\binom{-1/2}{k} = \frac{\sqrt{\pi}}{\Gamma(\frac{1}{2}-k)k!}$$

so that

$$(1+iq)^{-1/2} = \sqrt{\pi} \sum_{k=0}^{\infty} \frac{i^k}{\Gamma(\frac{1}{2}-k)k!} q^k$$

and Watson's lemma gives

$$\int_{\mathcal{C}_1^{(b)}} e^{i\lambda s^2} ds \simeq \sqrt{\pi} \sum_{k=0}^{\infty} \frac{i^k \Gamma(k+1)}{\Gamma(\frac{1}{2}-k)k!} \lambda^{-(k+1)}$$

Combining these results, noting that the integral over $\gamma = \mathcal{C} \cup \mathcal{C}'$ is zero by Cauchy's theorem,

$$\int_{\mathcal{C}} e^{i\lambda s^2} ds = 0 - \int_{\mathcal{C}'} e^{i\lambda s^2} ds \simeq \frac{(1+i)}{2\sqrt{\lambda}} \sqrt{\frac{\pi}{2}} - \sqrt{\pi} \sum_{k=0}^{\infty} \frac{i^k \Gamma(k+1)}{\Gamma(\frac{1}{2}-k)k!} \lambda^{-(k+1)} \quad (\lambda \to \infty).$$

In this example it was possible to globally parameterize the deformed contour and to derive a full asymptotic expansion. This might not always be possible. Since it is the local behaviour near the saddlepoint that dominates the integral, it can be shown that good approximations result by replacing the integral over the constant phase contour of steepest descent through the saddlepoint by a local approximation to this contour. In practice, any contour that crosses a saddlepoint and descends on both sides of it will typically produce a useful analytic approximation. The quality of the approximation so obtained should be checked numerically.

Exercise 17 . Consider the integral

$$I(\lambda) \triangleq \int_0^1 \log t e^{i\lambda t} dt \quad (\lambda \to \infty).$$

Why is the method of stationary phase not appropriate? Approximate the integral by the method of steepest descents.

ASYMPTOTIC METHODS AND STATISTICAL APPLICATIONS

Chapter 3: saddlepoint approximations in statistics

Heather Battey March 6, 2020

3.1 Introduction

Central to the construction of confidence sets and p-values is a specification of what would happen in hypothetical repeated application. Exact distribution theory for a test statistic or estimator of interest under hypothetical replication, is rarely available, and we resort to asymptotic arguments including the central limit theorem. We will derive corrections to the central limit theorem in inverse powers of the sample size, giving more refined distribution theory for an asymptotically standard normally distributed statistic, suitable for use in small samples. Some of these ideas generalize to statistics whose limiting distribution is not normal.

We will discuss the saddlepoint approximation (Daniels, 1954) to the density function of the mean of independent and identically distributed scalar-valued random variables whose cumulant generating function is known. This situation is very restrictive but the ideas involved generalize usefully.

3.2 Steepest descents derivation of the saddlepoint approximation

3.2.1 Problem specification

Let $f_X(x)$ denote the density function of a scalar-valued random variable X at x and let $M_X(t) = \mathbb{E}(e^{tX})$ for $t \in \mathbb{R}$ be the moment generating function of X. The latter might only exist in a relatively small region because the integral defining the expectation need not converge for all t. On the other hand, the characteristic function $\psi_X(t) = M_X(it)$ exists for all $t \in \mathbb{R}$ because $|e^{i\theta}| = |\cos(\theta) + i\sin(\theta)| = 1$ for all $\theta \in \mathbb{R}$. The functions M_X and f_X are related by Fourier inversion as

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} M_X(it)e^{-itx}dt.$$

Let $-c_1 < t < c_2$ be the interval, containing the origin, in which $\int e^{tx} f_X(x)dx$ converges. If a transform

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$$

converges for all real valued t in the range $-c_1 < t < c_2$, then its analytic continuation $M_X(s)$ must converge for all complex values s in the strip $-c_1 < \mathrm{Re}(s) < c_2$ (see e.g. Widder, 1941, p238). In other words, all the singularities of $M_X(s)$, when viewed as a function over the complex plane, are outside the strip $-c_1 < \mathrm{Re}(s) < c_2$ containing the imaginary axis. Modulo a sign change, $M_X(s)$ for $s \in \mathbb{C}$ is a bilateral Laplace transform of $f(x)$, given by

$$M_X(s) = \int_{-\infty}^{\infty} e^{sx} f(x)dx = \lim_{R\to\infty} \int_0^R e^{sx} f(x)dx + \lim_{R\to\infty} \int_{-R}^0 e^{sx} f(x)dx.$$

Introduce $X_1, \ldots, X_n$, independent copies of X, and define $Z_n = n^{-1} \sum_{i=1}^n X_i$. By direct calculation, the characteristic function of Z_n is $\{M_X(it/n)\}^n$ and

$$f_{Z_n}(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \{M_X(it/n)\}^n e^{-itz} dt, \quad z \in \mathbb{R}, \tag{3.2.1}$$

or, changing variables to $q = it/n$,

$$f_{Z_n}(z) = \frac{n}{2\pi i} \int_{-i\infty}^{i\infty} \exp\{n \log M_X(q)\} \exp\{-nqz\} dq = \frac{n}{2\pi i} \int_{-i\infty}^{i\infty} \exp[n\{K_X(q) - qz\}] dq \tag{3.2.2}$$

where, for real-valued argument, $K_X(t) = \log M_X(t)$ is the cumulant generating function of X. In (3.2.2), K_X takes a purely imaginary argument.

The density function of the sample mean has thus been expressed as an integral of the form $I_C(\lambda)$ of the last chapter, where $\lambda = n$ and the limiting operation is $n \to \infty$. The central limit theorem gives the limiting form of the density of $(\sqrt{n}/\sigma)Z_n$ as $n \to \infty$. Our asymptotic integral approximations of the last chapter are asymptotic forms rather than limits, and will typically give useful approximations for small n, thereby providing small sample refinements to the central limit theorem.

3.2.2 The Bromwich integral

Let $C_0 \triangleq (iy : -\infty < y < \infty)$, i.e., a line up the imaginary axis. Define $C_0(R) \triangleq (iy : -R < y < R)$ so that $\lim_{R\to\infty} C_0(R) = C_0$ and

$$f_{Z_n}(z) = \lim_{R\to\infty} \frac{n}{2\pi i} \int_{C_0(R)} \exp[n\{K_X(q) - qz\}] dq. \tag{3.2.3}$$

The contour $C_0(R)$ can be deformed to a new contour $C_{\text{new}}(R)$ without changing the integral (3.2.3), provided that $C_0(R) \cup C_{\text{new}}(R)$ is a closed curve, $\gamma(R)$ say, and that the interior of $\gamma(R)$ does not contain any singularities of $K_X(q) - qz$. This is guaranteed if the new contour does not go outside the region $-c_1 < \text{Re}(s) < c_2$. Consider two cases. If $C_{\text{new}}(R)$ is composed of a line $C_2'(R)$ running parallel to the *right* of $C_0(R)$ and two horizontal lines $C_1'(R)$ and $C_3'(R)$ joining it to $C_0(R)$, we have, writing $m_n(q) \triangleq \exp[n\{K_X(q) - qz\}]$,

$$\int_{C_0(R)} m_n(q)dq = \oint_{\gamma'(R)} m_n(q)dq - \int_{C_1'(R)} m_n(q)dq - \int_{C_2'(R)} m_n(q)dq - \int_{C_3'(R)} m_n(q)dq.$$

If $C_{\text{new}}(R)$ is composed of a line $C_2''(R)$ running parallel to the *left* of $C_0(R)$ and two horizontal lines $C_1''(R)$ and $C_3''(R)$ joining it to $C_0(R)$, we have

$$\int_{C_0(R)} m_n(q)dq = \oint_{\gamma''(R)} m_n(q)dq - \int_{C_1''(R)} m_n(q)dq - \int_{C_2''(R)} m_n(q)dq - \int_{C_3''(R)} m_n(q)dq.$$

It can be shown in either of these cases that the integrals parallel to the real axis tend to zero as $R \to \infty$. So, noting the direction of integration around the corresponding closed curves, represented by

$$\oint_{\gamma'(R)} \quad \text{and} \quad \oint_{\gamma''(R)},$$

equation (3.2.2) can be equivalently written as

$$f_{Z_n}(z) = \frac{n}{2\pi i} \int_{\tau - i\infty}^{\tau + i\infty} \exp[n\{K_X(s) - sz\}]ds, \tag{3.2.4}$$

where τ is anywhere in $(-c_1, c_2)$. This is the so-called Bromwich integral for inversion of Laplace transforms. Daniels (1954) obtained a simple asymptotic approximation to the integral defining $f_{Z_n}(z)$ by choosing τ to pass through a saddlepoint of the integrand in such a way that the integrand is negligible outside its immediate neighbourhood.

3.2.3 Application of the method of steepest descents

The saddlepoint-defining equation of Chapter 2 is $K_X'(s) = z$, an arbitrary solution to which is denoted by s^*. Daniels (1954) proved the following theorem.

Theorem 3.2.1. Define a and b to be such that $F_X(x) = 0$ for all $x < a$ and $F_X(x) = 1$ for all $x > b$, where F_X is the distribution function of X. Then for every $a < z < b$, there exists at least one root of $K_X'(s) = z$, denoted by $s^* = t^* + iy^*$, among which there is a unique real-valued root.

To obtain a simple analytic approximation to $f_{Z_n}(z)$, consider the real root of the saddlepoint-defining equation, which satisfies $s^* = t^*$. The local landscape around any saddlepoint is qualitatively similar to that of Figure 3 in Chapter 2, i.e. there are two constant phase paths through a saddlepoint. On one of which, the altitude function $u(t, y) \triangleq \mathrm{Re}(K_X(s) - sz)$ decays on either side of the saddlepoint, this is the steepest descent path. On the other path the altitude function increases on either side of the saddlepoint, this is the steepest ascent path. Care must be taken to ensure the descent path is used, otherwise the resulting approximation will be poor.

The phase function is $v(t, y) \triangleq \mathrm{Im}(K_X(s) - sz)$. At the saddlepoint $s^* = t^*$, $v(t^*, y^*) = v(t^*, 0) = 0$ so that, on constant phase paths through $s^* = t^*$, $K_X(s) - sz$ must be real. Because $K_X(t)$ is real for $t \in \mathbb{R}$, one constant phase path through $s^* = t^*$ is the real axis, and this is necessarily a steepest ascent or steepest decent path through $s^* = t^*$ by the Cauchy Riemann equations. When viewed as a function of a real variable t, $K_X(t) - tz$ achieves a *minimum* at t^*, which can be shown by checking that $K''_X(t^*) > 0$, a proof of which is given below. It follows that the real axis is a steepest *ascent* path through the saddlepoint and should be discarded. The steepest descent path is the other contour passing through the saddlepoint on which $\mathrm{Im}(K_X(s) - sz) = 0$. So on the steepest descent path, $K_X(s) - sz$ is a real-valued function of a complex argument.

Proof that $K''_X(t^*) > 0$. By definition $K_X(t) = \log M_X(t)$ so $K'_X(t) = \{M'_X(t)/M_X(t)\} = z$ at $t = t^*$ and

$$K''_X(t) = -\left\{\frac{M'_X(t)}{M_X(t)}\right\}^2 + \frac{M''_X(t)}{M_X(t)} = -z^2 + \frac{M''_X(t)}{M_X(t)}, \quad \text{at } t = t^*. \tag{3.2.5}$$

Define the function

$$M_X(t; z) = e^{K_X(t) - tz} = M_X(t)e^{-tz} = \int_{-\infty}^{\infty} e^{t(x-z)} f_X(x)dx.$$

Differentiating with respect to t gives

$$\begin{aligned} M'_X(t; z) &= \int_{-\infty}^{\infty} (x - z)e^{t(x-z)} f_X(x)dx \\ M''_X(t; z) &= \int_{-\infty}^{\infty} (x - z)^2 e^{t(x-z)} f_X(x)dx > 0. \end{aligned}$$

Similarly $M'_X(t) = \int xe^{tx} f_X(x)dx$ and $M''_X(t) = \int x^2 e^{tx} f_X(x)dx$ so that, expanding the expression for $M''_X(t; z)$ we have

$$\frac{M''_X(t; z)}{M_X(t, z)} = \frac{e^{-tz}M''_X(t) + e^{-tz}z^2 M_X(t) - 2e^{-tz}zM'_X(t)}{e^{-tz}M_X(t)} = \frac{M''_X(t)}{M_X(t)} + z^2 - 2z\frac{M'_X(t)}{M_X(t)}.$$

At $t = t^*$, this simplifies to

$$\frac{M''_X(t^*;z)}{M_X(t^*;z)} = \frac{M''_X(t^*)}{M_X(t^*)} - z^2.$$

Rearranging and substituting in (3.2.5), we see that

$$K''_X(t^*) = -z^2 + z^2 + \frac{M''_X(t^*;z)}{M_X(t^*;z)} > 0.$$

□

In any specific problem, K_X will be known and the contour $\{s : \mathrm{Im}(K_X(s) - sz) = 0\}$ that is not the real axis could be identified, ideally parameterized, and joined to $(s = t^* + iy : -\infty < y < y)$, say, by bypass paths along which the integral is negligible for large n leading to a full asymptotic expansion along this contour. Daniels (1954) instead obtained an asymptotic approximation on the contour $(s = t^* + iy : -\infty < y < y)$, arguing initially that on *any* admissible straight line parallel to the imaginary axis, the integrand $\exp[n\{K_X(s) - sz\}]$ attains its maximum modulus only where the line crosses the real axis, and away from this point, viewing n as fixed for the purpose of this argument, the integrand decays like $O(|y|^{-n})$ as $|y| \to \infty$. He subsequently shows that the approximation so obtained is identical to that which would result from a local parameterization of the constant phase contour of steepest descent through $s^* = t^*$.

A Taylor series expansion around $s^* = t^*$ on the line $s = t^* + iy$, $(-\infty < y < \infty)$ gives

$$K_X(s) - sz = K_X(s^*) - s^*z + \tfrac{1}{2}K''_X(s^*)\underbrace{(s - s^*)}_{iy}{}^2 + \sum_{k=3}^{\infty} \frac{K_X^{(k)}(s^*)}{k!}\underbrace{(s - s^*)}_{iy}{}^k,$$

or, changing variables to $v = \{nK''_X(s^*)\}^{1/2}y$,

$$K_X(s) - sz = K_X(t^*) - t^*z + \frac{v^2}{2n} - \frac{1}{3!}\frac{K'''_X(t^*)}{\{K''_X(t^*)\}^{3/2}}\frac{iv^3}{n^{3/2}} + \frac{1}{4!}\frac{K_X^{(4)}(t^*)}{\{K''_X(t^*)\}^2}\frac{v^4}{n} + \cdots, \qquad (3.2.6)$$

and writing in terms $\lambda_j(s) \triangleq K_X^{(j)}(s)/\{K''_X(s)\}^{j/2}$, we have

$$\begin{aligned} f_{Z_n}(z) &= \frac{n}{2\pi}e^{n\{K_X(t^*)-t^*z\}}\int_{-\infty}^{\infty}\exp\{-nK''_X(t^*)y^2/2\}\exp\left\{n\sum_{k=3}^{\infty}\frac{K_X^{(k)}(t^*)(-iy)^k}{k!}\right\}dy \\ &= \left\{\frac{n}{K''_X(t^*)}\right\}^{1/2}\frac{e^{n\{K_X(t^*)-t^*z\}}}{2\pi}\int_{-\infty}^{\infty}e^{-v^2/2}\left\{1 - \frac{\lambda_3(t^*)iv^3}{3!\sqrt{n}}\right. \\ &\qquad\qquad \left. + \frac{\lambda_4(t^*)v^4}{4!n} - \frac{\lambda_3^2(t^*)v^6}{(3!)^2n} + \cdots\right\}dv \end{aligned}$$

where the last line follows from equation (3.2.6), a Taylor series expansion of e^x around $x = 0$ and arranging in fractional powers of n^{-1}. The leading integral term is the Gaussian integral and evaluates to $\sqrt{2\pi}$, terms involving odd powers of v vanish on integration, and the remaining integrals can be evaluated exactly by changing variables to express them as gamma integrals. For instance, letting $t = v^2/2$,

$$\begin{aligned}\int_{-\infty}^{\infty} e^{-v^2/2} v^4 dv &= 2^{3/2}\int_0^{\infty} e^{-t} t^{2-1/2} dt = 4\sqrt{2}\int_0^{\infty} e^{-t} t^{5/2-1} dt = 4\sqrt{2}\Gamma(5/2)\\ \int_{-\infty}^{\infty} e^{-v^2/2} v^6 dv &= 2^3\sqrt{2}\int_0^{\infty} e^{-t} t^{7/2-1} dt = 2^3\sqrt{2}\Gamma(7/2).\end{aligned}$$

Since $\Gamma(z+1) = z\Gamma(z)$ and $\Gamma(1/2) = \sqrt{\pi}$, so that $\Gamma(5/2) = 3\sqrt{\pi}/4$ and $\Gamma(7/2) = 15\sqrt{\pi}/8$, there results the following approximation

$$f_{Z_n}(z) \simeq g_n(z)[1 + n^{-1}\{\lambda_4(t^*)/8 - 5\lambda_3^2(t^*)/24\} + \cdots], \tag{3.2.7}$$

where

$$g_n(z) = n^{1/2}\{2\pi K_X''(t^*)\}^{-1/2} \exp[n\{K_X(t^*) - t^* z\}]$$

is called the *saddlepoint approximation* to $f_{Z_n}(z)$.

3.3 Discussion

Small-sample refinements to the central limit theorem can also be obtained by Edgeworth expansions, which we will not derive in detail. The key idea is the following.

Let Z_n be a statistic that converges in distribution to the random variable Z as n tends to infinity. For instance, in the simplest setting, Z_n would be a standardized sum of i.i.d. random variables and Z would be a standard normally distributed random variable. Let f_{Z_n} and f_Z denote the density functions of these statistics and let $\kappa_{n,j}$ and ν_j denote the jth cumulants associated with each. Furthermore, let K_{Z_n} denote the cumulant generating function associated with Z_n. The Fourier transform of f_{Z_n} is

$$f_{Z_n}^*(t) = \exp\{K_{Z_n}(it)\} = \exp\Big\{\sum_{j=1}^{\infty}(it)^j\kappa_{n,j}/j!\Big\} = \exp\Big\{\sum_{j=1}^{\infty}(it)^j(\kappa_{n,j} - \nu_j)/j!\Big\} f_Z^*(t)$$

where f_Z^* is the Fourier transform of f_Z. Thus we have expressed the Fourier transform of f_{Z_n} in terms of that of the limiting density and small-sample corrections to the cumulants. The remaining step is to invert the Fourier transform. The resulting approximation to f_{Z_n} is called the Edgeworth expansion when f_Z is the standard normal density function ϕ and when terms are arranged in increasing powers of $n^{-1/2}$, so that the expansion is of the form

$$f_{Z_n}(z) = \phi(z)\{1 + a_{1,n}g_1(z) + a_{2,n}g_2(z) + \cdots\}. \tag{3.3.1}$$

The arrangement of the terms ensures that $a_{j,n} \to 0$ and $a_{j+1,n}/a_{j,n} \to 0$ as $n \to \infty$ so that the series (3.3.1) is asymptotic.

In the special case in which f_Z is the standard normal density, the inversion results in an expression in terms of Hermite polynomials, which are orthogonal with respect to the standard normal density function. This is coincidental rather than by construction, and gives the correction terms a useful interpretation.

The Edgeworth expansion only requires knowledge or sufficiently accurate estimation of the first few moments of Z_n, by contrast the saddlepoint approximation requires knowledge of the whole cumulant generating function of the original random variables. The expansion of (3.2.7) is an expansion in inverse powers of n rather than in inverse powers of $n^{1/2}$. Unlike in (3.3.1), the terms inside $\{\cdots\}$ in (3.2.7) do not depend on the evaluation point z. The relative error of the saddlepoint approximation is thus uniformly bounded by terms of order $O(n^{-1})$. Because the Hermite polynomials evaluated at z are unbounded as $|z| \to$, the Edgeworth expansion will tend to perform poorly in the tail areas, which are the areas of primary interest in calculating p-values, for instance.

These notes emphasized a derivation of the saddlepoint approximation in which the method of steepest descents is applied to the inverse Laplace transform of the density function. There is a second derivation based on an Edgeworth expansion applied after suitable exponential tilting. Both have advantages. The latter approach is usually emphasized in statistics books covering this topic. For a discussion of Edgeworth and Saddlepoint approximations in the context of multivariate and conditional distributions, see Barndorff-Nielsen and Cox (1979).

The following table gives transformations to $0/0$ and ∞/∞ for indeterminate forms.

limits of $f(x)$ and $g(x)$	indeterminate form	transformation to $0/0$	transformation to ∞/∞
$\lim_{x\to x_0} f(x) = 0$, $\lim_{x\to x_0} g(x) = 0$	$0/0$	–	$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{1/g(x)}{1/f(x)}$
$\lim_{x\to x_0} f(x) = \infty$, $\lim_{x\to x_0} g(x) = \infty$	∞/∞	$\lim_{x\to x_0} \frac{f(x)}{g(x)} = \lim_{x\to x_0} \frac{1/g(x)}{1/f(x)}$	–
$\lim_{x\to x_0} f(x) = 0$, $\lim_{x\to x_0} g(x) = \infty$	$0 \times \infty$	$\lim_{x\to x_0} f(x)g(x) = \lim_{x\to x_0} \frac{f(x)}{1/g(x)}$	$\lim_{x\to x_0} f(x)g(x) = \lim_{x\to x_0} \frac{g(x)}{1/f(x)}$
$\lim_{x\to x_0} f(x) = \infty$, $\lim_{x\to x_0} g(x) = \infty$	$\infty - \infty$	$\lim_{x\to x_0} \{f(x) - g(x)\} = \lim_{x\to x_0} \frac{1/g(x) - 1/f(x)}{1/f(x)g(x)}$	$\lim_{x\to x_0} \{f(x) - g(x)\} = \log \lim_{x\to x_0} \frac{\exp\{f(x)\}}{\exp\{g(x)\}}$
$\lim_{x\to x_0} f(x) = 0$, $\lim_{x\to x_0} g(x) = 0$	0^0	$\lim_{x\to x_0} f(x)^{g(x)} = \exp\left\{\lim_{x\to x_0} \frac{g(x)}{1/\log f(x)}\right\}$	$\lim_{x\to x_0} f(x)^{g(x)} = \exp\left\{\lim_{x\to x_0} \frac{\log f(x)}{1/g(x)}\right\}$
$\lim_{x\to x_0} f(x) = 1$, $\lim_{x\to x_0} g(x) = \infty$	1^∞	$\lim_{x\to x_0} f(x)^{g(x)} = \exp\left\{\lim_{x\to x_0} \frac{\log f(x)}{1/g(x)}\right\}$	$\lim_{x\to x_0} f(x)^{g(x)} = \exp\left\{\lim_{x\to x_0} \frac{g(x)}{1/\log f(x)}\right\}$
$\lim_{x\to x_0} f(x) = \infty$, $\lim_{x\to x_0} g(x) = 0$	∞^0	$\lim_{x\to x_0} f(x)^{g(x)} = \exp\left\{\lim_{x\to x_0} \frac{g(x)}{1/\log f(x)}\right\}$	$\lim_{x\to x_0} f(x)^{g(x)} = \exp\left\{\lim_{x\to x_0} \frac{\log f(x)}{1/g(x)}\right\}$

Table A.1: Table of indeterminate forms for application of l'Hôpital's rule.

APPENDIX B: PRINCIPLE TAYLOR SERIES EXPANSIONS

The five principle Taylor series expansions around zero are

$$
\begin{aligned}
e^x &= 1 + x + \tfrac{x^2}{2!} + \cdots + &&= \sum_{n=0}^{\infty} \frac{x^n}{n!}, \quad (-\infty < x < \infty);\\
\sin x &= x - \tfrac{x^3}{3!} + \tfrac{x^5}{5!} - \cdots + &&= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}, \quad (-\infty < x < \infty);\\
\cos x &= 1 - \tfrac{x^2}{2!} + \tfrac{x^4}{4!} - \cdots + &&= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}, \quad (-\infty < x < \infty);\\
(1+x)^m &= 1 + mx + \tfrac{m(m-1)x^2}{2!} + \cdots + &&= \sum_{n=0}^{\infty} \frac{m(m-1)\cdots(m-n+1)x^n}{n!}\\
& &&= \sum_{n=0}^{\infty} \binom{m}{n} x^n, \quad (-1 < x < 1);\\
\log(1+x) &= x - \tfrac{x^2}{2} + \tfrac{x^3}{3} - \cdots + &&= \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n}, \quad (-1 < x < 1).
\end{aligned}
$$

From these and the formula for the sum of a geometric progression, $\sum_{k=0}^{\infty} ar^k = a/(1-r)$, $|r| < 1$, expansions in powers of x for many other functions are simply obtained.

Solution to Exercise 1 . Assuming that all derivatives are uniformly bounded away from zero and upper bounded by B we have the upper bound (1.2.2) or, equivalently,

$$\begin{aligned} |R_K(x)| &\leq \frac{B}{(K+1)K!}|x-x_0|^K|x-x_0| \\ &= \frac{f^{(K)}(x_0)|x-x_0|^K}{k!}\frac{B|x-x_0|}{(K+1)|f^{(K)}(x_0)|}. \end{aligned}$$

The first of these terms is $|f_K(x)|$ and we see that

$$\frac{|R_K(x)|}{|f_K(x)|} \leq \frac{B|x-x_0|}{(K+1)|f^{(K)}(x_0)|} \to 0 \qquad (x \to x_0).$$

Solution to Exercise 2 . Part (a): base functions $\psi_k(x) = x^k$. The first few coefficients a_k of the expansion $\tan(x) = \sum_{k=1}^{\infty} a_k\psi_k(x)$ are

$$\begin{aligned} a_0 &= \lim_{x\to 0}\frac{tanx}{1} = 0 \\ a_1 &= \lim_{x\to 0}\frac{tanx}{x} = \lim_{x\to 0}\frac{\sec^2 x}{1} = 1 \\ a_2 &= \lim_{x\to 0}\frac{tanx - x}{x^2} = \lim_{x\to 0}\frac{\sec^2 x - 1}{2x} = \lim_{x\to 0}\frac{2\sec^2 x\tan x}{2} = 0 \\ a_3 &= \lim_{x\to 0}\frac{tanx - x}{x^3} = \lim_{x\to 0}\frac{\sec^2 x - 1}{3x^2} = \lim_{x\to 0}\frac{2\sec^2 x\tan x}{6x} = \lim_{x\to 0}\frac{2\sec^4 x + 2\tan x(d/dx)\sec^2 x}{6} = \frac{1}{3}. \end{aligned}$$

All even terms are seen to be zero and we obtain

$$\tan x = x + \frac{x^3}{3} + O(x^5) \quad x \to 0.$$

Part (b): base functions $\psi_k(x) = \sin^k(x)$.

$$\begin{aligned} a_0 &= \lim_{x\to 0}\frac{tanx}{\sin^0 x} = 0 \\ a_1 &= \lim_{x\to 0}\frac{tanx}{\sin x} = \lim_{x\to 0}\frac{\sec^2 x}{\cos x} = 1 \\ a_2 &= \lim_{x\to 0}\frac{tanx - \sin x}{\sin^2 x} = \lim_{x\to 0}\frac{\sec^2 x - \cos x}{2\sin x\cos x} = \lim_{x\to 0}\frac{2\sec^2 x\tan x + \sin x}{2\cos^2 x - 2\sin^2 x} = 0 \\ a_3 &= \lim_{x\to 0}\frac{tanx - \sin x}{\sin^3 x} = \lim_{x\to 0}\frac{\sec^2 x - \cos x}{3(d/dx)\sin x(\sin^2 x)} \\ &= \lim_{x\to 0}\frac{2\sec^2 x\tan x + \sin x}{-3\sin^2 x + 6\cos^2 x\sin x} = \lim_{x\to 0}\frac{2\sec^4 x + 4\tan^2 x\sec^2 x + \cos x}{-9\cos x\sin^2 x - 12\sin^2 x\cos x + 6\cos^3 x} = \frac{1}{2}. \end{aligned}$$

Again, all even terms are seen to be zero and we obtain

$$\tan x = \sin x + \frac{\sin^3 x}{2} + O(\sin^5 x) \quad x \to 0.$$

Solution to Exercise 3 . By symmetry $\mathbb{E}(R) = 0$ so that

$$\mathrm{var}(R) = \frac{\Gamma(p/2)}{\sqrt{\pi}\Gamma\{(p-1)/2\}} \int_{-1}^{1} r^2(1-r^2)^{(p-3)/2} dr = \frac{2\Gamma(p/2)}{\sqrt{\pi}\Gamma\{(p-1)/2\}} \int_{0}^{1} r^2(1-r^2)^{(p-3)/2} dr$$

Changing variables to $s = r^2$ we have

$$\begin{aligned} \mathrm{var}(R) &= \frac{2\Gamma(\frac{p}{2})}{2\sqrt{\pi}\Gamma(\frac{p-1}{2})} \int_0^1 s^{1-1/2}(1-s)^{(p-3)/2} ds \\ &= \frac{\Gamma(p/2)B(\frac{3}{2},\frac{p-1}{2})}{\sqrt{\pi}\Gamma(\frac{p-1}{2})} = \frac{\Gamma(\frac{p}{2})}{2\Gamma(\frac{p}{2}+1)} \simeq \frac{1}{2}\frac{2}{p} = p^{-1}. \end{aligned}$$

Solution to Exercise 4 . For any $x \in (0,1)$,

$$\mathrm{pr}(M > x) = \mathrm{pr}(\cap_{i=1}^{k-1}\{V_i > x\}) = \{1 - F_V(x)\}^{k-1} = (1-x)^{k-1}.$$

Therefore

$$\begin{aligned} \mathrm{pr}(M > X) &= \int_0^1 \mathrm{pr}(M > x) f_X(x) dx = (1-\gamma)\int_0^1 (1-x)^{k-1} x^{-\gamma} dx \\ &= (1-\gamma)\frac{\Gamma(1-\gamma)\Gamma(k)}{\Gamma(1-\gamma+k)} \simeq k^{-(1-\gamma)} \to k^{-1} \quad (\gamma \to 0). \end{aligned}$$

For the statistical motivation for this problem, see Cox and Battey (2017) and especially Battey and Cox (2018).

Solution to Exercise 5 . For $n = 0, 1, 2, \ldots$ we have

$$\int_{-\infty}^{\infty} e^{-tx^2} x^{2n} dx = 2\int_0^{\infty} e^{-tx^2} x^{2n} dx = t^{-1/2} t^{-n} \int_0^{\infty} e^{-s} s^{n-1/2} ds = \frac{\Gamma(n+1/2)}{t^{n+1/2}}.$$

Solution to Exercise 6 . By Stirling's formula we have

$$\frac{\Gamma(\frac{n}{2}+\frac{1}{2})}{\Gamma(\frac{n}{2})} \simeq \left(\frac{n}{2}\right)^{1/2} \quad (n \to \infty).$$

so that the constant term in f_X is approximately $(2\pi)^{1/2}$ for large n. In $(1+x^2/n)^{-(n+1)/2}$, the sample size n appears as a power. Taking logs and expanding for small x^2/n we have

$$\begin{aligned} -\frac{(n+1)}{2}\log\left(1+\frac{x^2}{n}\right) &= -\frac{(n+1)}{2}\left\{\sum_{k=1}^{\infty}(-1)^{k+1}\frac{(x^2/n)^k}{k}\right\} \\ &= -\frac{x^2}{2} + n^{-1}\left(\frac{x^4}{4} - \frac{x^2}{2}\right) + n^{-1}\left(\frac{x^4}{4} - \frac{x^6}{6}\right) + O(n^{-3}). \end{aligned}$$

So that

$$f_X(x) = \phi(x)\left\{1 + n^{-1}\left(\frac{x^4}{4} - \frac{x^2}{2}\right) + n^{-1}\left(\frac{x^4}{4} - \frac{x^6}{6}\right)\right\} + O(n^{-3})$$

where ϕ is the standard normal density function.

Solution to Exercise 7 . The motivation for this questions comes from Battey and Cox (2018). The notation used in the question corresponds to that in the paper but to simplify presentation here, write $X \triangleq V_{S2}$. The conditional probability generating function of X given $V_{S1} = v_{S1}$ is

$$\begin{aligned}\mathbb{E}_{X|V_{S1}=v_{S1}}(\eta^X) &= \sum_{x=0}^{v_{S1}} \eta^x \binom{v_{S1}}{x} \theta_{S2.1}^x (1-\theta_{S2.1})^{v_{S1}-x} \\ &= (1-\theta_{S2.1}+\theta_{S2.1}\eta)^{v_{S1}}\end{aligned}$$

by the binomial theorem. The unconditional probability generating function is therefore, by the binomial theorem again,

$$\begin{aligned}\mathbb{E}_X(\eta^X) &= \sum_{v_{S1}=0}^{v_{S0}} (1-\theta_{S2.1}+\theta_{S2.1}\eta)^{v_{S1}} \binom{v_{S0}}{v_{S1}} \theta_{S1}^{v_{S1}} (1-\theta_{S1})^{v_{S0}-v_{S1}} \\ &= (1-\theta_{S1}\theta_{S2.1}+\theta_{S1}\theta_{S2.1}\eta)^{v_{S0}},\end{aligned}$$

showing that V_{S2} is a binomially distributed random variable with index v_{S0} and parameter $\theta_{S1}\theta_{S2.1}$.

Solution to Exercise 8 . Write $X \triangleq V_{N2}$. As before $\mathbb{E}_{X|V_{N1}=v_{N1}}(\eta^X) = (1-\theta_{N2.1}+\theta_{N2.1}\eta)^{v_{N1}}$, so that unconditionally

$$\begin{aligned}\mathbb{E}_X(\eta^X) &= \sum_{v_{N1}=0}^{\infty} (1-\theta_{N2.1}+\theta_{N2.1}\eta)^{v_{N1}} \frac{e^{-\mu_{N1}}\mu_{N1}^{v_{N1}}}{v_{N1}!} \\ &= e^{-\mu_{N1}}e^{\mu_{N1}}\exp(-\mu_{N1}\theta_{N2.1}+\mu_{N1}\theta_{N2.1}\eta),\end{aligned}$$

the probability generating function of a Poisson random variable with parameter $\mu_{N1}\theta_{N2.1}$.

Solution to Exercise 9 . The function f_X is the unstandardized Gumbel density function. The standardized version is one of the three types of extreme value limit laws for the distribution of standardized maxima or minima. See Fisher and Tippett (1928) for the original derivation and Gnedenko (1943) for a formal proof of the so called *three types theorem.*

The moment generating function is, with $s = e^{\alpha/\beta}e^{-x/\beta}$,

$$M(t) = \int_{-\infty}^{\infty} dx e^{tx} w(x) = -\beta^{-1}e^{t\alpha}\int_{\infty}^{0} s^{-t\beta}(\beta/s)se^{-s}ds = e^{t\alpha}\Gamma(1-t\beta),$$

so that

$$\begin{aligned}
\mathrm{I}_1 &\triangleq -\Gamma'(1)\beta + \alpha = \gamma\beta + \alpha, \\
\mathrm{I}_2 &\triangleq \Gamma''(1)\beta^2 + 2\gamma\beta\alpha + \alpha^2 \\
\mathrm{I}_3 &\triangleq -\Gamma'''(1)\beta^3 + 3\Gamma''(1)\beta^2\alpha - 3\Gamma'(1)\beta\alpha^2 + \alpha^3 \\
\vdots & \qquad\qquad \vdots \\
\mathrm{I}_m &\triangleq \sum_{k=0}^{m} \binom{m}{k} (-1)^k \Gamma^{(k)}(1)\beta^k\alpha^{m-k},
\end{aligned}$$

where γ is Euler's constant.

The derivatives of the gamma function obey the recursion $\Gamma^{(m)}(z) = \Gamma(z)R_m(z)$, where

$$R_m(z) = R'_{m-1}(z) + \psi(z)R_{m-1}(z),$$

$R_1(z) = \psi(z) = (d/dz)\log\Gamma(z)$ is the digamma function, satisfying (Abramowitz and Stegun, 1964, 6.4.10)

$$\psi^{(m)}(z) = (-1)^{m+1} m! \sum_{k=0}^{\infty} (z+k)^{-(m+1)}.$$

Evaluating at $z = 1$ gives the sum $\sum_{k=1}^{\infty} k^{-(m+1)} = \zeta(m+1)$, where ζ is Riemann's zeta function (Abramowitz and Stegun, 1964, 23.2.18). Thus

$$\begin{aligned}
\Gamma'(1) &= -\gamma, \qquad\qquad\qquad \Gamma''(1) = \pi^2/6 + \gamma^2, \\
\Gamma^{(3)}(1) &= -\{2\zeta(3) + \gamma^3 + \gamma\pi^2/2\}, \quad \Gamma^{(4)}(1) = \gamma^4 + \gamma^2\pi^2 + \frac{3\pi^4}{20} + 8\gamma\zeta(3) \\
\Gamma^{(5)}(1) &= -\left\{\gamma^5 + \frac{5\gamma^3\pi^2}{3} + \frac{3\gamma\pi^4}{4} + 24\zeta(5) + \frac{10}{3}(6\gamma^2 + \pi^2)\zeta(3)\right\} \\
\Gamma^{(6)}(1) &= \gamma^6 + \frac{5\gamma^4\pi^2}{2} + \frac{9\gamma^2\pi^4}{4} + \frac{61\pi^6}{168} + 20\gamma(2\gamma^2 + \pi^2)\zeta(3) + 40\{\zeta(3)\}^2 + 144\gamma\zeta(5),
\end{aligned}$$

where we have used equations 23.2.24 and 23.2.25 of Abramowitz and Stegun (1964) in the last three calculations.

Solution to Exercise 10 .

$$I(\lambda) = \int_0^\infty e^{-\lambda t}\cos t\,dt, \quad \lambda \to \infty.$$

Expanding $\cos t$,

$$I(\lambda) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} \underbrace{\int_0^\infty e^{-\lambda t} t^{2n} dt}_{I_n(\lambda)}$$

where $I_n(\lambda)$ can be solved exactly by making a change of variable $s = \lambda t$ giving

$$I_n(\lambda) = \lambda^{-(2n+1)}\Gamma(2n+1) = \lambda^{-(2n+1)}(2n)!.$$

$$I(\lambda) = \sum_{n=0}^{\infty} \frac{(-1)^n}{\lambda^{2n+1}} = \frac{\lambda}{1+\lambda^2},$$

the same answer as by integration by parts.

Solution to Exercise 11 . The incomplete gamma function is given by

$$\gamma_m(z) = \frac{1}{\Gamma(m)} \int_z^\infty x^{m-1} e^{-x} dy = \frac{1}{\Gamma(m)} \int_z^\infty \underbrace{x^{m-1}}_{u} \underbrace{d(-e^{-x})}_{dv}$$

and integration by parts gives

$$\begin{aligned} \gamma_m(z) &= \left[-\frac{x^{m-1}e^{-x}}{\Gamma(m)}\right]_z^\infty + \frac{(m-1)}{\Gamma(m)} \int_z^\infty y^{m-2} e^{-y} dy \\ &= \frac{z^{m-1}e^{-z}}{\Gamma(m)} \left\{1 + \frac{(m-1)}{z} + \frac{(m-1)(m-2)}{z^2} + \cdots \right\}. \end{aligned}$$

Solution to Exercise 12 . The integral is not in a form to which Watson's lemma can be applied. Try changing variables to $u = \cosh t$. This is monotonically increasing on $(0, \infty)$ and minimized at $t = 0$ at which $\cosh(0) = 1$. The limits of integration for u are thus 1 and ∞, so this is still not of a form to which Watson's lemma applies. Consider instead $s = \cosh t - 1$. We have

$$\int_0^\infty \frac{e^{-\lambda \cosh t}}{\sin^{1/3} t} dt = e^{-\lambda} \int_0^\infty \frac{e^{-\lambda(\cosh t - 1)}}{\sin^{1/3} t} dt.$$

$$dt = \frac{1}{\sqrt{s(s+2)}} ds \simeq \frac{1}{\sqrt{2s}} \quad (t \to 0),$$

and by the small angle formula $\sin^{1/3} t \simeq t^{1/3}$ $(t \to 0)$. Expanding $\cosh t$ for small t in the equation $\cosh t - 1 = s$ shows that

$$\frac{t^2}{2!} + \frac{t^4}{4!} + \cdots = s$$

or

$$\frac{t^2}{2!} + o(t^2) = s, \quad (t \to 0)$$

so that $t \simeq \sqrt{2s}$ $(t \to 0)$. We have, since $t \to 0$ corresponds to $s \to 0$,

$$\int_0^\infty \frac{e^{-\lambda \cosh t}}{\sin^{1/3} t} dt \simeq e^{-\lambda} \int_0^\infty \frac{e^{\lambda s}}{(2s)^{1/6}(2s)^{1/2}} = \frac{e^{-\lambda}}{2^{2/3}} \underbrace{\int_0^\infty \frac{e^{-\lambda s}}{s^{2/3}} ds}_{I(\lambda)}.$$

This is of the form $\int g(s) e^{-\lambda s} ds$ with $g(s) \simeq s^{-2/3}$ $(s \to 0)$. Watson's lemma applies, giving

$$I(\lambda) = \lambda^{-1/3}\Gamma(1/3) + O(\lambda^{-1/3}).$$

Solution to Exercise 13 . Write

$$\{F(u+r/2)-F(u-r/2)\}^{n-2} = \exp[(n-2)\log\{F(u+r/2)-F(u-r/2)\}].$$

The integral in the density function of the sample range is now of Laplace type with $\lambda = (n-2)$.

This approach to approximating the density function of the sample range is due to Cox (1948).

Solution to Exercise 14 . The gamma function is $\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt$. Change variables to $t = xs$ so that $dt = xds$,

$$e^{-t}t^{x-1} = exp\{(x-1)\log(xs) - xs\} = x^{x-1}s^{-1}\exp\{x(\log s - s)\}$$

and

$$\Gamma(x) = x^x \int_0^\infty s^{-1}\exp\{x(\log s - s)\}ds$$

Apply Laplace's method for large x with $u(s) = \log s - s$ giving $u'(s) = s^{-1} - 1$, $u''(s) = -s^{-2}$. This is minimised at $c = 1$, at which $u(1) = -1$, $u''(1) = 1$ and the Laplace approximation to $\Gamma(x)$ is

$$\Gamma(x) \simeq x^x \sqrt{\frac{2\pi}{x}} e^{-x} \quad (x \to \infty).$$

Solution to Exercise 15 . In the notation of §2.4, $u(t) = \sin t$, $u'(t) = \cos t$, $u''(t) = -\sin t$. There is a maximum at $\pi/2$ at which $u(\pi/2) = 1$ and $u''(\pi/2) = -1$. So the Laplace approximation to the integral is

$$\int_0^\pi e^{x\sin t}dt \simeq \sqrt{\frac{2\pi}{x}} e^x \quad (x \to \infty).$$

Solution to Exercise 16 . (i) $u(t) \simeq u(a) + u'(a)(t-a)$ $(t \to a)$. The contribution outside an interval $[a, a+\varepsilon]$ can be shown to be exponentially small as in the proof of Watson's lemma and we have

$$\begin{aligned} I_L(\lambda) &\simeq g(a)e^{\lambda u(a)} \int_a^{a+\varepsilon} e^{\lambda u'(a)(t-a)}dt \\ &= \frac{g(a)e^{\lambda u(a)}}{-\lambda u'(a)} \int_0^{-\lambda u'(a)\varepsilon} e^{-s}ds \simeq \frac{g(a)e^{\lambda u(a)}}{-\lambda u'(a)} \underbrace{\int_0^\infty e^{-s}ds}_{=1} \qquad (\lambda \to \infty). \end{aligned}$$

(ii) If $u'(a) = 0$ and $u''(a) < 0$ the same strategy leads us to

$$I_L(\lambda) \simeq \frac{g(a)e^{\lambda u(a)}}{\{-\lambda u''(a)\}^{1/2}} \int_0^{\{-\lambda u''(a)\}^{1/2}\varepsilon} e^{-s^2/2}ds \simeq \frac{g(a)e^{\lambda u(a)}}{\{-\lambda u''(a)\}^{1/2}} \underbrace{\int_0^{\{-\lambda u''(a)\}^{1/2}\varepsilon} e^{-s^2/2}ds}_{=\sqrt{\pi/2}} \qquad (\lambda \to \infty).$$

Solution to Exercise 17 . The method of stationary phase is not appropriate here because there are no stationary points of the phase function so the main contribution comes from the boundary. In simpler cases, such a situation can be dealt with by integration by parts but this approach is invalidated here because $\log t$ explodes at the boundary.

Framing the problem in terms of the notation of section 2.6, the altitude function is identically zero and the phase function is $v(t,0) = t$. The phases at the endpoints are zero and one and the constant phase paths through the end points are therefore $C_v^0 \triangleq \{(t,y) : v(t,y) = 0\}$ and $C_v^1 \triangleq \{(t,y) : v(t,y) = 1\}$. This suggests deforming the original integration path $\mathcal{C} \triangleq \{(t,0) : t \in [0,1]\}$ to $\mathcal{C}' = \mathcal{C}_1^{(T)} \cup \mathcal{C}_2^{(T)} \cup \mathcal{C}_3^{(T)}$, where for some value of T

$$\begin{aligned} \mathcal{C}_1^{(T)} &= \{(0,y) : y \in [0,T]\}, \\ \mathcal{C}_2^{(T)} &= \{(t,T) : t \in [0,1]\}, \\ \mathcal{C}_3^{(T)} &= \{(1,y) : y \in [T,0]\}. \end{aligned}$$

This choice means that we take the two constant phase paths (straight lines) emanating from the end points and connect them where the contribution to the integral is small for large λ for for a suitable choice of T. Consider

$$\int_{\mathcal{C}_2^{(T)}} \log s e^{i\lambda s} ds = e^{-\lambda T} \int_0^1 \log(t + iT) e^{i\lambda t} dt.$$

This tends to zero as $T \to \infty$, while the integral over $\mathcal{C}_1^{(T)}$ is analytically tractable by a change of variables and that over $\mathcal{C}_3^{(T)}$ can be approximated by Watson's lemma.

REFERENCES

Abramowitz, M. and Stegun, I. (1972), *Handbook of Mathematical Functions*, Dover, NY.

Barndorff-Nielsen, O. and Cox, D. R. (1979), Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. R. Statist. Soc, B.*, 41, 279–312.

Battey, H. S. and Cox, D. R. (2018), Large numbers of explanatory variables: a probabilistic assessment. *Proc. R. Soc. Lond. A.*, 747.

Bender, C. and Orzag, S. (1999). *Asymptotic Methods and Perturbation Theory.* Springer, NY.

Comtet, L. (1974) Advanced Combinatorics, Reidel Publishing, Dordrecht.

Cox, D. R. (1948). A note on the asymptotic distribution of the range. *Biometrika*, 35, 310–315.

Cox, D. R. (1988). Some aspects of conditional and asymptotic inference: A review. *Sankhya A*, 50, 314–337.

Cox, D. R. and Battey, H. S. (2017) Large numbers of explanatory variables, a semi-descriptive analysis. *Proc. Nat. Acad. Sci.*, 114, 8592–8595.

Daniels, H. E. (1954), Saddlepoint approximations in statistics, *Ann. Math. Statist*, 25, 631–650.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10, 507–521.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest of a sample. *Proc. Camb. Phil. Soc*, 24, 180–190.

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Ann. Math.*, 44, 423–453.

McCullagh, P. (2017), *Tensor Methods in Statistics*, Dover, New York.

Watson, G. N. (1944). *A Treatise on the Theory of Bessel Functions*, 2nd edition, Cambridge University Press.

Widder, D. V. (1941), *The Laplace Transform.* Princeton University Press, Princeton, NY.