607
Asymptotic Expansions: Some basics

Jon Faust

http://e105.org/e607

October 13, 2016

- **More reading**
  - On expansions, Rothenberg's Handbook of Econometrics chapter on this topic
    go
    http://www.sciencedirect.com/science/article/pii/S1573441284020079
  - On the bootstrap, Hansen's bootstrap chapter in his web text and the Wright notes p.7–11.
  - Hall's book, The bootstrap and Edgeworth Expansion, is a very nice more general treatment treatment
  - See also, all the cites in the bootstrap section of the reading list
- **First order asymptotics**
  - We have a very nice set of tools based on conventional asymptotics
  - But commonly these techniques don't perform well
  - That is, the approximation error when using standard asymptotic methods is often large for DGPs and sample sizes relevant in macro.
  - Put another way: Any asymptotically justified method actually suggests a large family of methods that only differ in finite samples. In practical cases, some members of the family may provide decent approximations, others provide poor approximations. And it is really difficult to know which to use in any given case.
- **Bootstrap: motivation**
  - The bootstrap as we'll use it, can be seen as a method for picking among asymptotically equivalent methods

- In particular, we'll write an expansion (a bit like a Taylor series approximation) for the distribution of the statistics in which
- Conventional asymptotics will be seen as exploiting 1 term in the expansion (a first order approximation)
- Asymptotically equivalent methods are those that have the same first-order term.
- There are various ways to effectively do a second or third order approximation.
- Thus, we might pick among equivalent methods based on the higher terms.

  Using higher-order terms is called performing an 'asymptotic refinement'

- These ideas have been around for over a century.
- This all sounds great, except
- Until recently this was almost never done in practice because exploiting the higher order terms was very complicated

  and the benefits just didn't seem to outweigh the complexity costs.

- Around 1980, folks discovered that we would get these asymptotic refinement benefits using the bootstrap.
- And bootstrapping is really, really easy. And it seems to work well across a broad range of cases.
- This lecture sketches some technical background on asymptotic expansions as background for more concretely discussing bootstrap virtues.
- We'll focus only on Edgeworth expansions

▶ Background: Expansions

- We all know Taylor series expansions:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)(x - x_0)^n}{n!}$$

  where $f^{(n)}$ is, well, you know...

- and power series expansions:

$$e^x = \sum x^n/n! = 1 + x + x^2/2! + \dots$$

▶ Stochastic expansions

- Edgeworth expansions are a form of expansion of the distribution function of a statistic

▶ **Edgeworth expansions**

- Suppose some statistic ϕ_T has distribution function F_T.
- Let's just posit that:

$$F_T(c) \approx g_0(c) + g_1(c)/T^{1/2} + g_2(c)/T^1 + \ldots + g_J(c)/T^{j/2}$$

that is,

$$F_T(c) \approx \sum_{j=0}^{J} \frac{g_j(c)}{T^{j/2}}$$

- This expression means nothing or is entirely vacuous until we've said something about the approximation error.

  We'll do that in a moment, but to understand what we'll do, you first need to think about the expression

- We have expanded not one function (F), but a sequence of functions, F_T, $T = 1, \ldots \infty$
- The left-hand side varies with T, but T only enters the RHS in the denominators.

  The gs are fixed functions, and don't vary with T.

▶ **For intuition, fix** c

- Fix c and the gs become just a set of J numbers.
- The equation then says that we are approximating $F_T(c)$ for every T by taking a weighted combination of J fixed numbers g.
- And only the weights on the gs change with T (and do so in a simple way).
- Obviously, the contribution of each g_j after g_0 falls with T.
- The larger is J the faster g_j's contribution falls
- So if the approximation works in a sensible way, the higher order terms are picking up features of F that disappear as the sample size grows.
- Put another way, in a really big sample, perhaps you just need the first-order term to get a reasonable approx.
- In somewhat smaller, but still large samples, 2 terms might be needed to attain that quality
- In even smaller, but still large, samples, 3 terms ...
- And so forth.

▶ **Sadly**

- We do reason in precisely this way

- But as we'll see these are very strange approximations and so the merits of going to higher order terms are more questionable than, say, in the case of Taylor series or power series approximations.

▶ **Error in the approximation**

- As we said, until we talk about how the error term in the approximation varies, our equation defining the approximation is vacuous.
- To do this, it is useful to define some new notation.
- Notationally, a (nonstochastic) sequence a_T that vanishes (goes to zero) is muliplied by T^p is said to be of order $o(T^{-p})$

  one says, 'little o T to the minus p'

- You could read $o(T^{-p})$ as goes to zero like (or at least as fast as) T^{-p}
- Note also that if we add any fixed, finite number of terms that are $o(T^{-p})$, the result is of the same order

  each of the finite number of terms goes to zero, so the sum does as well.

- For a nice discussion and its generalization to stochastic sequences, see, Mc Fadden's notes McFadden's Ch. 4 Notes

  http://elsa.berkeley.edu/users/mcfadden/e240a_sp01/ch4.pdf

▶ **Back to the expansion**

- Our expansion:

$$F_T(c) \approx \sum_{j=0}^{J} \frac{g_j(c)}{T^{j/2}}$$

- Suppose we add that the approximation error (which must be a function of T) vanishes faster with T than the last included term

$$F_T(c) = \sum_{j=0}^{J} \frac{g_j(c)}{T^{j/2}} + o(T^{-J/2})$$

  Note we now have an equal sign: that last term is whatever it takes to justify the equal sign, and what we know about it is how fast it goes to zero.

- Let's add that this statement about the error is uniform in c
- Now we have an Edgeworth expansion

▶ **Edgeworth expansion**

- Take a statistic ϕ_T with distribution function F_T.
- Suppose

$$F_T(c) = \sum_{j=0}^{J} g_j(c)/T^{j/2} + err_T(c)$$

  and $err_T(c)$ is $o(T^{-J/2}$ uniformly in c.
- That is,

$$\sup_c |T^{J/2} err_T(c)| \to 0$$

- When these conditions are met, this is an Edgeworth expansion for F_T.

▶ **Existence**

- You might ask whether for F_T of interest, g_js that satisfy this condition tend to exist
- The answer is that often they do not exist

  Many statistics do not have a valid Edgeworth expansion.

- Only when functions g_j with the stated property exist, do we say that the statistic has a valid $J + 1$-term Edgeworth expansion.

  For emphasis: Edgeworth expansions of any given order need not exist in any given case. That is, there may be a 1 term expansion, but no two term expansion.

▶ **Comments**

- Thus, if we have derived a conventional asymptotic distribution for a statistic, we know that a 1 term Edgeworth expansion exists
- And $g_0(c)$ must be the conventional asymptotic distribution
- So for statistics that are asymptotically normal, g_0 is the Gaussian distribution function.
- If the statistic is asymptotically pivotal, g_0 doesn't depend on any unknown parameters.

  by definition of pivotal

- Generally, all gs involve parameters of the problem

  That is, most statistics are not asymptotically pivotal.

- And even when the statistic is asymptotically pivotal, g_js for $j > 0$ involve the parameters of the problem.

  That is, our pivotal notion seldom gets us much further than the initial term. (There are some exceptions to this.)

- We can now view conventional asymptotics as using the first term of the expansion

- This gives rise to the label first order asymptotics—taking only the first term of the expansion.
- One way to refine our approaches (that is, to pick among the family of statistics with the same first order term) is to consider higher order terms

  Called asymptotic refinement.

▶ **Some important details**

- Unlike most every type of expansions you are familiar with, there is no guarantee that the Edgeworth expansion converges

  quite generally it does not

- That is, suppose an infinite-order expansion exists. The infinite sum need not converge!

  Fix a T and a c and the sum of the gs need not converge.

- That is a bit of a mind-bending fact. Nonconvergent series are odd beasts in mathematics.
- Essentially for the same reason, adding one term to your expansion need not make the approximation more accurate for a given fixed T and c

  Thus, properly capturing a higher order term need not deliver a more accurate approximation.

- For any given T, J, and c, the J term approximation need not even be in the unit interval

  Since F is a distributions function, its a pretty good guess that $F(c)$ is in the unit interval.

- Nonetheless, we often proceed as if taking advantage of one more term is a good thing and, hence, deserves a nice-sounding label like 'refinement'
- At one time folks did very laborious explicit derivations of expansions for purposes like that just stated.

  And these derivations get very laborious very quickly.

- This stuff was very messy and did not reliably lead to significantly better results in practical cases and this kind of work basically died.
- Rothenbergs Handbook of Ecmet chapter covers this material.

  go

  http://www.sciencedirect.com/science/article/pii/S1573441284020079

- Basic summary is: these derivations are cumbersome and this work flourished for a time, but largely tailed off without leaving much of an imprint on practical work

▶ **But recently . . .**

- More recently bootstrap techniques have been shown to capture or take account of higher order terms in asymptotic expansions.
- And we can get the benefits using the Monte Carlo/resampling methods described in previous lectures without ever explicitly deriving the expansions

  Our techniques only require that a valid expansion exists, but we don't need to know it.

- The discovery of this justification for the bootstrap then led to an immense shift toward bootstrap techniques.
- I'll sketch that argument in the next lecture.

▶ **Finally, note**

- 1.There are related expansions and concepts you'll quickly come across in the theoretical literature:

  Cornish-Fisher expansions, Barry-Essen Bounds, Saddlepoint approximations; see the sources noted at the beginning.

- 2. If you dive into the theory literature, you'll find most of the discussion in terms of cumulants of distributions.

  If you are not familiar with cumulants and wish to be so, see the next few comments

▶ **Aside:: Cumulants**

- If you decide to venture into the theory underlying the expansions we are talking about you will quickly begin to read about cumulants.
- If you don't know what these are, they will perplex you a bit.

  That is, you may understand the definition, but wonder how and why people came to define such things.

- This is a famously baffling topic. If you google `intutive explanation of cumulants`, you mainly find failed attempts. Or attempts that take you quickly into discussions of Feynman and quantum mechanics.
- This is good and bad. You should get the correct impression that
  - cumulants are quite fundamental in many contexts, and
  - perpliexing nonetheless.

▶ **Aside::**

- I can't help much here, but I can give you an analogy that may help
- You should understand moments and their relation to moment generating functions (MGFs)

- And you remember that the moment generating function is one way to define distributions.

▶ **Aside::**

- We are often interested in understanding the distribution of (normalized) sums of random variables and how it is related to the underlying distribution of the items.
- Here the MGF is not too much help: there is not trivial route from from the MGFs of $X_1, X_2, \ldots$ to that of $X_1 + X_2 + \ldots$.
- Suppose I wanted something like an analogous function to the MGF, call it a CGF, for which the CGF of a sum is just the sum of the underlying CGFs
- Then I could conveniently analyze the distribution of sums in terms of the underlying distributions.
- The cumulant generating function (CGF) has this property.
- That is, until you have a better sense, think of the CGF as something analogous to the MGF but which has the virtue that the CGF of a sum is the sum of the CGFs.

▶ **Aside:: Fun facts: CGFs and the CLT**

- Note: For the $N(0, \sigma^2)$ distribution, the cumulants are $0, \sigma^2, 0, \ldots$.

  Hmmm. Normals must be special.

- Standard central limit theorems show that normalized sums converge to normals.
- This must mean the the sum of the cumulants of underlying random variables must all converge to zero, except the second.

▶ **Finally**

- A lovely (in my view) introduction to cumulants is provided by the first few introductory pages of:

  Three lectures on free probability,Jonathan Novak go

  http://arxiv.org/pdf/1205.2097v1.pdf

- After introduction, this paper it goes off into the weird world of something called 'free probability,' which I know essentially nothing about. I'm not suggesting you read that, but the first couple sections motivating cumulants are quite nice.