# LECTURE 11: CONFIDENCE SETS

Michal Kolesár*

September 30, 2024

REFERENCE This is not really covered in HMC, apart from some basics in Chapter 4.2. Chapter 9 in CB is a much better reference. H22, Chapter 14 and 16.13.

## 1. INTRODUCTION

In this lecture, we will study confidence sets. Let $X$ denote our data. Let $\theta \in \mathbb{R}$ be our parameter of interest. Our task is to construct a data-dependent set $C(X)$ that contains $\theta$ with large probability. In standard problems, the confidence set for a scalar parameter will often take form of an interval, $[L(X), U(X)]$. One possibility is to set $L(X) = -\infty$ and $U(X) = \infty$. Such an interval will contain $\theta$ with probability 1. Of course, the problem with this interval is that it is too long. So, we want to construct an interval that is as short as possible (or, if $\dim(\theta) > 1$, a set with the smallest possible volume), subject to a coverage requirement.

Let us introduce a couple of basic concepts related to confidence sets.

*Definition 1.* The coverage probability of the set $C(X) \subset \Theta$ at $\theta$ is given by $P_\theta(\theta \in C(X))$.

Of course, in practice, we are interested in confidence sets that contain the true parameter value with large probability uniformly over the set of possible parameters values.

*Definition 2.* The confidence set $C(X)$ has confidence level $1-\alpha$ if $\inf_{\theta\in\Theta} P_\theta(\theta \in C(X)) \geq 1-\alpha$.

Note the randomness comes from the data $X$ so that the set $C(X)$ changes in repeated sampling. Let us consider how we can construct confidence sets.

## 2. TEST INVERSION

For each possible parameter value $\theta_0 \in \Theta$, consider the problem of testing $H_0 \colon \theta = \theta_0$ against $H_1 \colon \theta \neq \theta_0$. Suppose that for each such hypothesis we have a test $\phi_{\theta_0}(X)$ of level

*Email: mkolesar@princeton.edu.

$\alpha$. Then the confidence set

$$C(X) = \{\theta_0 \in \Theta : \; H_0 \colon \theta = \theta_0 \text{ is not rejected}\} = \{\theta \in \Theta \colon \phi_\theta(X) = 0\}$$

has confidence level $1-\alpha$. Indeed, suppose that the true value of parameter is $\theta_0$. Since the test $\phi_{\theta_0}$ has level $\alpha$, by construction, $E_{\theta_0}[\phi_{\theta_0}(X)] = P_{\theta_0}(\phi_{\theta_0}(X) = 1) \le \alpha$. So, for any $\theta$,

$$P_\theta(\theta \in C(X)) = P_\theta(\phi_\theta(x) = 0) \ge 1 - \alpha,$$

which implies $\inf_{\theta\in\Theta} P_\theta\{\theta \in C(X)\} \ge 1-\alpha$. This procedure is called *test inversion*. One problem with the test inversion is that sometimes a confidence set obtained via this procedure will consist of several disjoint intervals which is unattractive from a practitioner's perspective.

Conversely, if we have a way to construct a confidence set with confidence level $1-\alpha$, we can use it for testing $H_0 \colon \theta = \theta_0$ against a two-sided alternative by letting $\phi_{\theta_0}(X) = 1$ if $\theta_0 \notin C(X)$. In other words, we accept the null if $\theta_0 \in C(X)$. This test will have level $\alpha$.

*Example 1.* Let $X_1, \dots, X_n$ be a random sample from distribution $\mathcal{N}(\theta, \sigma^2)$, $\sigma^2$ known. Let us use test inversion to construct a confidence set for $\theta$ of level $1-\alpha$. Consider the problem of testing the null hypothesis, $H_0 \colon \theta = \theta_0$ against the alternative, $H_1 \colon \theta \neq \theta_0$. The uniformly most powerful (UMP) unbiased test rejects if $|\overline{X}_n - \theta_0| \ge z_{1-\alpha/2}\sigma/\sqrt{n}$. Thus, $\theta \in C(X)$ iff

$$-z_{1-\alpha/2}\sigma/\sqrt{n} \le \overline{X}_n - \theta \le z_{1-\alpha/2}\sigma/\sqrt{n},$$

or, rewriting this, the confidence set is given by

$$\left[\overline{X}_n - z_{1-\alpha/2}\sigma/\sqrt{n} \;\; \overline{X}_n + z_{1-\alpha/2}\sigma/\sqrt{n}\right] \tag{1}$$

So in this case, we actually get an interval, ⊠

*Example 2.* Let $X_1, \dots, X_n$ be a random sample from the distribution $\mathcal{N}(\mu, \sigma^2)$. Let us use the test inversion to construct a confidence set for $\sigma^2$ of level $1-\alpha$. Consider the problem of testing the null hypothesis, $H_0 \colon \sigma^2 = \sigma_0^2$ against the alternative, $H_1 \colon \sigma^2 \neq \sigma_0^2$. Under the null hypothesis, $(n-1)S_n^2/\sigma_0^2 \sim \chi^2(n-1)$. The test that accepts the null hypothesis if and only if

$$\chi^{-2}_{\alpha/2}(n-1) \le (n-1)S_n^2/\sigma_0^2 \le \chi^{-2}_{1-\alpha/2}(n-1).$$

has size $\alpha$. This test will accept the null hypothesis $\sigma^2 = \sigma_0^2$ if and only if

$$(n-1)S_n^2/\chi^{-2}_{1-\alpha/2}(n-1) \le \sigma_0^2 \le (n-1)S_n^2/\chi^{-2}_{\alpha/2}(n-1),$$

so again, the confidence set is an interval, which has coverage $1-\alpha$ independently of the value of $\mu$. Notice the quantiles $\chi^{-2}_{\alpha/2}(n-1)$ and $\chi^{-2}_{1-\alpha/2}(n-1)$ get switched. ⊠

In general, if we can find a pivotal quantity $Q = q(X, \theta_0)$ such that distribution of $Q$

under the null hypothesis $\theta = \theta_0$ does not depend on the choice of $\theta_0$ or any other nuisance parameters, then we can use $Q$ for testing and confidence set construction. In particular, since the distribution of $Q$ is independent of the true parameter value, we can find numbers $a$ and $b$ such that $P_\theta\{a \le q(X, \theta) \le b\} = 1 - \alpha$ for all $\theta \in \Theta$. Then we can construct a level $\alpha$ test by accepting the null hypothesis that $\theta = \theta_0$ if and only if $a \le q(X, \theta_0) \le b$. The confidence set will consist of all parameter values $\theta_0$ which are accepted:

$$C(X) = \{\theta \in \Theta \colon a \le q(X, \theta) \le b\}.$$

## 3. PRATT'S THEOREM (OPTIONAL)

Intuitively, inverting powerful tests should yield short confidence intervals (confidence sets with small volume). Inverting tests that are UMP against a particular alternative $\theta_1$ should yield confidence intervals that are as short as possible when $\theta_1$ is the true parameter. The next result formalizes this intuition.

*Theorem 3 (Pratt 1961). Let $X \sim f(x \mid \theta)$, let $C(X)$ be a confidence set for $\theta \in \mathbb{R}$ based on inverting tests $\phi_\theta(X)$, and let $\lambda(C(X)) = \int_\theta (1 - \phi_\theta(X)) d\theta$ denote its volume. Then, for any $\theta_1$,*

$$E_{\theta_1}[\lambda(C(X))] = \int_\Theta (1 - \beta_\theta(\theta_1))\, \mathrm{d}\theta = \int_\Theta P_{\theta_1}(\theta \in C(X))\, \mathrm{d}\theta,$$

*where $\beta_\theta(\theta_1) = E_{\theta_1}[\phi_\theta]$ denotes the power of the test $\phi_\theta$ against $\theta_1$. Furthermore, if the tests $\phi_\theta(X)$ are UMP against $\theta_1$, and the expected length is finite, then $C(X)$ achieves the shortest expected length at $\theta_1$ among all confidence sets with level $1 - \alpha$.*

*Proof.* The first result follows from changing the order of integration:

$$E_{\theta_1}[\lambda(C(X))] = E_{\theta_1}\left[\int_\Theta (1 - \phi_\theta(X))\, \mathrm{d}\theta\right] = \int_\Theta \int_{\mathcal{X}} (1 - \phi_\theta(x)) f(x \mid \theta_1)\, \mathrm{d}x\, \mathrm{d}\theta = \int_\Theta (1 - \beta_\theta(\theta_1))\, \mathrm{d}\theta.$$

If $C(X)$ is based on inverting tests $\phi_\theta$ that are UMP against $\theta_1$, then for each $\theta \neq \theta_1$, such tests minimize $1 - \beta_\theta(\theta_1)$, and therefore also minimize the integral. □

The second part, in spite of often being cited to justify the practice of constructing confidence sets by inverting tests, is not that useful since it only tells us about how to minimize the length of confidence interval (CI) at a particular point, and often, this comes at the expense of terrible expected length at other points (generically, we won't simultaneously be able to minimize length at all $\theta$). A more useful result would tell us how to construct a CI that minimizes its maximum length $\sup_\theta E_\theta[\lambda(C(X))]$, or average expected length, averaged over $\theta$'s using some weights $w(\theta)$, or how one can ever justify reporting one-sided CIs (which have infinite length and therefore Pratt's result doesn't apply). We'll leave those theorems for later courses.

*Remark 4.* The current theorem is useful, however, as a benchmark for the best possible performance of a CI in more complicated problems: if we construct a CI that is "close", in some sense, to the Pratt bound uniformly over $\theta$, then we know one cannot improve much over such CI.

*Example 1 (continued).* Since CI in Example 1 is based on inverting UMP unbiased tests, the CI in eq. (1) is the CI with the shortest expected length among all intervals based on inverting unbiased tests. However, stated in this way, it's not clear why one should restrict attention to unbiased tests.

What if we don't restrict attention to unbiased tests? Then we can use Theorem 3 as a benchmark to establish for the best possible performance of any CI at some given $\theta_1$. Specifically, by Theorem 3, the confidence set with the shortest expected length when $\theta = \theta_1$ is given by inverting UMP tests of the null $H_0\colon \theta = \theta_0$ against the alternative $H_1\colon \theta = \theta_1$. By the Neyman-Pearson lemma, these are one-sided $z$-tests that reject if $\overline{X}_n - \theta_0 \geq \zeta$ if $\theta_0 < \theta_1$, and reject if $\overline{X}_n - \theta_0 \leq -\zeta$ if $\theta_0 > \theta_1$, where $\zeta = z_{1-\alpha}\sigma/\sqrt{n}$. This means that the CI consists of all values of $\theta$ that either satisfy ($\overline{X}_n - \zeta < \theta$ and $\theta < \theta_1$), or else satisfy ($\overline{X}_n + \zeta \geq \theta$ and $\theta > \theta_1$). We may write this as

$$\{\theta\colon \overline{X}_n - \zeta < \theta < \theta_1 \quad \text{or} \quad \theta_1 \leq \theta \leq \overline{X}_n + \zeta\}$$

Based on the value of $\overline{X}_n$, it may be the case that either both sets of inequalities may hold at once, or else one of them cannot hold. Breaking down the possible cases, we obtain the CI

$$\begin{cases} [\overline{X}_n - \zeta, \theta_1] & \text{if } \overline{X}_n + \zeta \leq \theta_1, \\ [\overline{X}_n - \zeta, X + \zeta] & \text{if } \overline{X}_n - \zeta \leq \theta_1 \leq \overline{X}_n + \zeta, \\ [\theta_1, \overline{X}_n + \zeta] & \text{if } \overline{X}_n - \zeta \geq \theta_1, \end{cases} \tag{2}$$

It then follows by some algebra (fill it in!) that the expected length of the CI when $\theta = \theta_1$ is

$$\frac{2\sigma}{\sqrt{n}}\left[z_{1-\alpha}(1-2\alpha) + E[(Z + z_{1-\alpha})\mathbb{1}\{Z > z_{1-\alpha}\}]\right] = \frac{2\sigma}{\sqrt{n}}\left[z_{1-\alpha}(1-\alpha) + \varphi(z_{1-\alpha})\right].$$

where $Z$ is standard normal and $\varphi$ is the standard normal probability density function (PDF). In contrast, the length of the CI in eq. (1) is $2\sigma/\sqrt{n} \cdot z_{1-\alpha/2}$. Thus, when $\theta = \theta_1$, the efficiency of the CI in eq. (1) is

$$\frac{z_{1-\alpha}(1-\alpha) + \varphi(z_{1-\alpha})}{z_{1-\alpha/2}},$$

which equals 84.99% when $\alpha = 0.05$. Thus, the CI in eq. (1) is "nearly" efficient in the sense that it is not possible to improve upon in (in terms of expected length) by more than 15% even if we correctly "direct power" at the right $\theta$.

Note that the interval in eq. (2) can be arbitrarily long if $\overline{X}_n$ is far from $\theta_1$. So this interval is not something that we'd want to use in practice: we get a CI that is slightly shorter when $\overline{X}_n$ is close to $\theta_1$ at the expense of being very long when $\overline{X}_n$ is far from $\theta_1$. ⊠

# 4. ASYMPTOTIC THEORY FOR INTERVAL CONSTRUCTION

Let $X_1, \ldots, X_n$ be a random sample from distribution $f(x \mid \theta)$ with $\theta \in \Theta$. Under some regularity conditions,

$$\sqrt{n}(\hat{\theta}_{ML} - \theta) \Rightarrow N(0, \mathcal{I}^{-1}(\theta)).$$

For any function $g \colon \Theta \to \mathbb{R}$, under some regularity conditions, by delta-method

$$\sqrt{n}(g(\hat{\theta}_{ML}) - g(\theta)) \Rightarrow N(0, \nabla g(\theta)' \mathcal{I}^{-1}(\theta) \nabla g(\theta)).$$

We can consistently estimate the asymptotic variance $(g'(\theta))^2 \mathcal{I}^{-1}(\theta)$ by

$$\hat{V} = \nabla g(\hat{\theta}_{ML})' \left( -\nabla^2 \ell_n(\hat{\theta}_{ML})/n \right)^{-1} \nabla g(\hat{\theta}_{ML}).$$

Then, by Slutsky theorem,

$$\frac{\sqrt{n}(g(\hat{\theta}_{ML}) - g(\theta))}{\hat{V}^{1/2}} \Rightarrow \mathcal{N}(0,1),$$

so that we can construct an approximate CI for $g(\theta)$ as

$$g(\hat{\theta}_{ML}) \pm z_{1-\alpha/2} \hat{V}^{1/2} / \sqrt{n}.$$

The quantity $\hat{V}^{1/2}/\sqrt{n}$ is called the *standard error*.

Note that this confidence set is constructed based on inverting the second version of the Wald statistic that we considered in the previous lecture note,

$$\tilde{\zeta}_{Wald}(g_0) = \frac{n(g(\hat{\theta}_{ML}) - g_0)^2}{\hat{V}}.$$

*Digression (Uniform size control).* Because the Wald test $\phi_{g_0,Wald}(X_1, \ldots, X_n) = \mathbb{1}\{\tilde{\zeta}_{Wald} \geq z^2_{1-\alpha/2}\}$ controls size, so that

$$\lim_{n\to\infty} E_{\theta \colon g(\theta)=g_0}[\phi_{g_0,Wald}] \leq \alpha,$$

it follows that for each $\theta$, $\lim_{n\to\infty} P_\theta(g(\theta) \in C(X)) = 1 - \alpha$, and therefore that

$$\inf_\theta \lim_{n\to\infty} P_\theta(g(\theta) \in C(X)) = 1 - \alpha$$

(the CI achieves nominal coverage in large samples pointwise in $\theta$). However, it does not immediately follow that

$$\lim_{n\to\infty} \inf_{\theta\in\Theta} P_\theta(g(\theta) \in C(X)) = 1 - \alpha$$

(i.e. that for $n$ large enough, the level is approximately $1-\alpha$, this means that the CI achieves nominal coverage in large samples *uniformly* in $\theta$), although it's possible to show that under additional regularity conditions (which, however, fail in some important models). ⊠

*Example 3.* Let $X_1, \ldots, X_n$ be a random sample from Bernoulli$(p)$. Suppose we want to construct a confidence set for the odds ratio $g(p) = p/(1-p)$. We have $\hat{p}_{ML} = \overline{X}_n$, and

$$\sqrt{n}(\hat{p}_{ML} - p) \Rightarrow N(0, p(1-p)).$$

In addition,

$$g'(p) = \frac{(1-p)+p}{(1-p)^2} = \frac{1}{(1-p)^2}.$$

By the delta method,

$$\sqrt{n}(g(\hat{p}) - g(p)) \Rightarrow N(0, p/(1-p)^3).$$

So, $\hat{V} = \hat{p}_{ML}/(1-\hat{p}_{ML})^3$. Thus, an approximate confidence interval for $p/(1-p)$ is

$$\left[\frac{\hat{p}_{ML}}{1-\hat{p}_{ML}} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}_{ML}}{(1-\hat{p}_{ML})^3 n}}, \frac{\hat{p}_{ML}}{1-\hat{p}_{ML}} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}_{ML}}{(1-\hat{p}_{ML})^3 n}}\right]. \qquad \boxtimes$$

### 4.1. *Confidence Sets Based on LM and LR Tests*

In addition to the Wald statistic, we can invert tests based on the Lagrange multiplier (LM) and the likelihood ratio (LR) statistics as well. However, these confidence sets are usually more involved.

*Example 3 (continued).* Let's consider a CI for $p$. The CI based on inverting a Wald test is given by

$$\hat{p}_{ML} \pm z_{1-\alpha/2}\sqrt{\hat{p}_{ML}(1-\hat{p}_{ML})/n}.$$

To derive a CI based on inverting the LR test, note that the joint log-likelihood is

$$\ell_n(p) = \log\left(p^{\sum X_i}(1-p)^{n-\sum X_i}\right) = n\overline{X}_n \log p + n(1-\overline{X}_n)\log(1-p).$$

So

$$\begin{aligned}\zeta_{LR}(p) &= 2n\left[\overline{X}_n \log(\hat{p}_{ML}/p) + (1-\overline{X}_n)\log((1-\hat{p}_{ML})/(1-p))\right] \\ &= 2n\left[\overline{X}_n \log(\overline{X}_n/p) + (1-\overline{X}_n)\log((1-\overline{X}_n)/(1-p))\right]\end{aligned}$$

This yields the CI

$$\{p \in [0,1] \colon 2n\left[\overline{X}_n \log(\overline{X}_n/p) + (1-\overline{X}_n)\log((1-\overline{X}_n)/(1-p))\right] \leq z^2_{1-\alpha/2}\}$$

It is the solution to a nonlinear inequality.

To derive a CI based on inverting the LM test, note

$$S_n(p) = n\left(\frac{\overline{X}_n}{p} - \frac{1-\overline{X}_n}{1-p}\right) = n\frac{\overline{X}_n - p}{p(1-p)}, \qquad \mathcal{I}(p) = \frac{1}{p(1-p)}.$$

Thus,

$$\zeta_{LM}(p) = \frac{n(\overline{X}_n - p)^2}{p(1-p)}.$$

We know that $\zeta_{LM}(p) \Rightarrow \chi^2_1$. So, the confidence set based on inverting the $LM$ test is

$$\left\{p \in [0,1] \colon \left|\frac{\sqrt{n}(\overline{X} - p)}{\sqrt{p(1-p)}}\right| \leq z_{1-\alpha/2}\right\}.$$

It is the solution to a quadratic inequality. ⊠

## 5. BERNSTEIN-VON MISES

While the form of the confidence sets differs in small samples, one can show that CIs based on inverting Wald, LM and LR tests are all asymptotically equivalent. How do they compare to Bayesian credible intervals?

*Theorem 5. Suppose that the conditions for asymptotic normality of $\hat{\theta}_{ML}$ hold. Consider the rescaled and recentered posterior distribution for the parameter $\zeta = \sqrt{n}(\theta - \theta_0)$ under a continuous prior distribution $\Pi(\theta)$, such that the prior density $\pi(\theta)$ is continuous at $\theta_0$. Then the posterior distribution for $\zeta$ converges in probability to the distribution $\mathcal{N}(\sqrt{n}(\hat{\theta}_{ML} - \theta_0), \mathcal{I}(\theta_0)^{-1})$ (in total variation distance, that is $\sup_A |\Pi(\zeta \in A \mid X)| - P(A \in \mathcal{N}(\sqrt{n}(\hat{\theta}_{ML} - \theta_0), \mathcal{I}(\theta_0)^{-1})) \xrightarrow{p} 0$).*

*Proof.* See van der Vaart (1998, Theorem 10.1) for a rigorous proof. The parenthetical remark about total variation distance is needed because both the posterior distribution, and the normal approximating distributions are random: the former depends on the data, and the latter depends on $\hat{\theta}_{ML}$, which is random. To talk about two random distributions being close in probability, we need convert the difference between them into a single number, which is what the total variation distance does.

We prove the result pointwise, that is, we just look at the posterior evaluated at a single point $\zeta$. Now, by the change of variables formula, the posterior density for $\zeta$ is given by $\pi(\theta_0 + \zeta/\sqrt{n} \mid X)/\sqrt{n}$. If we scale the posterior by the posterior at $\theta_0$, we get

$$\frac{\pi(\theta_0 + \zeta/\sqrt{n} \mid X)}{\pi(\theta_0 \mid X)} = \frac{\mathcal{L}_n(\theta_0 + \zeta/\sqrt{n})\pi(\theta_0 + \zeta/\sqrt{n})}{\mathcal{L}_n(\theta_0)\pi(\theta_0)}.$$

Since $\pi$ is continuous at $\theta_0$, $\pi(\theta_0 + \zeta/\sqrt{n})/\pi(\theta_0) \to 1$. Furthermore, the log-likelihood difference satisfies, by a Taylor expansion,

$$\ell_n(\theta_0 + \zeta/\sqrt{n}) - \ell_n(\theta_0) = \zeta'\mathcal{S}_n(\theta_0)/\sqrt{n} + \frac{1}{2n}\zeta'\nabla^2\ell_n\zeta,$$

where each elements of the second derivative $\nabla^2 \ell_n$ are evaluated at intermediate values of $\theta$ that lie between $\theta_0$ and $\theta_0 + \zeta/\sqrt{n}$. By the usual hand-waving, this suggests

$$\ell_n(\theta_0 + \zeta/\sqrt{n}) - \ell_n(\theta_0) = \zeta' \mathcal{S}_n(\theta_0)/\sqrt{n} - \frac{1}{2n}\zeta' \mathcal{I}(\theta_0)^{-1}\zeta + \text{remainder terms.} \tag{3}$$

On the other hand, if we evaluate the normal distribution at a point $\eta$, the log-density is proportional to

$$-(\sqrt{n}(\hat{\theta}_{ML} - \theta_0) - \eta)' \mathcal{I}(\theta_0)(\sqrt{n}(\hat{\theta}_{ML} - \theta_0) - \eta).$$

In the proof of asymptotic normality in Lecture 6, we saw that up to remainder terms, $\sqrt{n}(\hat{\theta}_{ML} - \theta_0)$, equals $\mathcal{I}(\theta_0)^{-1} S_n(\theta_0)/\sqrt{n}$, so that, discarding things that don't depend on $\eta$, in large samples the log-density will equal eq. (3). □

This is a remarkable result. The first remarkable thing is that the limit distribution does not depend on the prior: the prior gets dominated by the likelihood, since as $n \to \infty$, whatever one's prior beliefs are, they get dominated by the data. Loosely speaking, we have that for $n$ large, the posterior for $\theta$ is given by

$$\theta \mid X \approx \mathcal{N}(\hat{\theta}, \mathcal{I}(\theta_0)^{-1}/n), \tag{4}$$

while in Lecture 6, we have proved that $\hat{\theta}_{ML} \mid \theta \approx \mathcal{N}(\theta, \mathcal{I}(\theta_0)^{-1}/n)$.

The second remarkable thing is that as a consequence of eq. (4), whatever Bayes decision one makes based on the posterior, so long as the decision is sufficiently smooth in the posterior, the Bayes decision will in large samples converge to a Bayes rule based on Equation (4). In particular:

1. Bayes posterior mean and posterior median estimators will in large samples be equivalent to $\hat{\theta}_{ML}$. They inherit all the large-sample optimality properties of maximum likelihood estimator (MLE).

2. Bayes credible sets will in large samples be equivalent to $\hat{\theta}_{ML} \pm z_{1-\alpha/2} \mathcal{I}(\theta_0)^1/n$. In particular, they will be equivalent to CIs based on inverting Wald, LM, or LR tests.

ONE PARAGRAPH SUMMARY OF THIS COURSE In Lecture 7, we showed that in finite-samples admissible rules are Bayes, and the reverse also. This result tells us that Bayes rules are optimal in large samples, and equivalent to likelihood-based inference. In regular parametric models, there is no wedge large samples, between optimal frequentist and Bayesian methods. But Bayesian methods have the additional advantage that they are optimal in finite samples in terms of average risk. Furthermore, in regular parametric models, estimation and inference just boil down to inference on $\theta$ based on a single normal observation $\mathcal{N}(\theta, \mathcal{I}(\theta_0)^{-1}/n)$.

## 6. (OPTIONAL) EXCITING NEW DEVELOPMENTS

This course has covered "standard" results in statistical theory. There are many recent results that are quite exciting, and that thing about statistics in a very different way. I will here three such results. If you manage to figure out how to make use of these in econometrics, I think you'll have a thesis topic sorted out!

1. E-values. For instance, doi.org/10.1214/23-STS894
2. The HulC https://doi.org/10.1093/jrsssb/qkad134 and universal inference https://doi.org/10.1073/pnas.1922664117
3. Conformal prediction.

## REFERENCES

Pratt, John W. 1961. "Length of Confidence Intervals." *Journal of the American Statistical Association* 56, no. 295 (September): 549–567. https://doi.org/10.2307/2282079.

Van der Vaart, Aad. 1998. *Asymptotic Statistics.* New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511802256.