Lecture 7: Frequentist Model Averaging

Francis J. DiTraglia

March 23, 2014

1 Hjort & Claeskens (2003)

Compromise Estimators: From Lemma 3.2, we know that

$$\hat{\delta}_S \equiv \sqrt{n}\left(\hat{\gamma}_S - \gamma_{0,S}\right) \xrightarrow{d} D_S = K_S \pi_S K^{-1}(\delta + W)$$

In the case of the full model, that is $S = \left\{1, 2, \xrightarrow{d} ots, q\right\}$ and $\pi_S = I_q$ so that $K_S = K$, this gives

$$D_n \equiv \hat{\delta}_{full} \equiv \sqrt{n}\left(\hat{\gamma}_{full} - \gamma_0\right) \xrightarrow{d} D = (\delta + W)$$

Thus, *any* submodel estimator $\hat{\delta}_S$ of δ converges in distribution to a linear combination of D, while the full model estimator of D_n of δ simply converges in distribution to D. In other words, the behavior of $\hat{\delta}_S$ is "essentially determined" by that of D_n. More precisely, the difference between $\hat{\delta}_S$ and $K_S \pi_S K^{-1} D_n$ is at most $o_p(1)$. Now consider a **Compromise Estimator** of the form:

$$\hat{\mu} = \sum_S c\left(S|D_n\right)\hat{\mu}_S$$

that is, we weight and sum submodel estimators where the weights are a function of $D_n = \hat{\delta}_{full} = \sqrt{n}(\hat{\gamma}_{full} - \gamma_0)$. *To ensure consistency, the weights must sum to one.*

Notation: Define G, a $q \times q$ matrix of functions, by

$$G(\xrightarrow{d} ot) = K^{-1/2} \left\{ \sum_S c(S| \xrightarrow{d} ot) H_S \right\} K^{1/2}$$

and $\hat{\delta}(D)$, an estimator of δ based on D, by

$$\hat{\delta}(D) = G(D)'D$$

Since H_S is symmetric and the weights $c(\xrightarrow{d} ot| \xrightarrow{d} ot)$ are scalars,

$$\hat{\delta}(D) = \left[K^{-1/2} \left\{ \sum_S c(S|D) H_S \right\} K^{1/2} \right]' D = K^{1/2} \left\{ \sum_S c(S|D) H_S \right\} K^{-1/2} D$$

Theorem 4.1 As long as the weight functions $c(\xrightarrow{d} ot| \xrightarrow{d} ot)$ sum to one and have at most a countable number of discontinuities, then

$$\sqrt{n}\,(\hat{\mu} - \mu_{true}) \xrightarrow{d} \sum_S c(S|D) \Lambda_S \equiv \Lambda$$

and

$$\Lambda = \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \omega' \left[\delta - \hat{\delta}(D) \right]$$

This is, in general, a **non-normal distribution** with

$$\begin{aligned} \text{mean} &= \omega' \left\{ \delta - E\left[\hat{\delta}(D) \right] \right\} \\ \text{variance} &= \tau_0^2 + \omega' Var\left[\hat{\delta}(D) \right] \omega \end{aligned}$$

where

$$\begin{aligned} \tau_0^2 &\equiv \nabla + \theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} \\ \omega &\equiv J_{10} J_{00}^{-1} \nabla_\theta \mu(\theta_0, \gamma_0) - \nabla_\gamma \mu(\theta_0, \gamma_0) \end{aligned}$$

and the MSE of Λ is

$$E[\Lambda^2] = \tau_0^2 + R(\delta)$$

where

$$R(\delta) = \omega' \left[\left\{ \hat{\delta}(D) - \delta \right\} \left\{ \hat{\delta}(D) - \delta \right\}' \right] \omega$$

Proof. First, using the fact that $\sum_S c(S|D_n) = 1$, we have

$$\begin{aligned} \sqrt{n}\left(\hat{\mu} - \mu_{true}\right) &= \sqrt{n}\left[\sum_S c(S|D_n)\hat{\mu}_S - \mu_{true}\right] \\ &= \sqrt{n}\left[\sum_S c(S|D_n)\hat{\mu}_S - \left\{\sum_S c(S|D_n)\right\}\mu_{true}\right] \\ &= \sum_S \left[c(S|D_n)\sqrt{n}\left(\hat{\mu}_S - \mu_{true}\right)\right] \end{aligned}$$

So we see that $\sqrt{n}\left(\hat{\mu} - \mu_{true}\right)$ is an almost-surely continuous function of the submodel estimators $\sqrt{n}\left(\hat{\mu}_S - \mu_{true}\right)$ and $D_n = \sqrt{n}\left(\hat{\gamma}_{full} - \gamma_0\right)$. Thus, to find the limiting distribution of the compromise estimator, we can apply the continuous mapping theorem, provided we have *joint* convergence in distribution of the submodel estimators and D_n.

Fortunately, we have already established precisely this joint convergence! In Lemma 3.3, we showed that the limit distribution of each submodel estimator $\sqrt{n}\left(\hat{\mu}_S - \mu_{true}\right)$ is a linear combination of

$$\begin{pmatrix} M \\ N \end{pmatrix} \sim \mathcal{N}_{p+q}(0, J_{full})$$

Further the limit distribution of $D_n = \sqrt{n}\left(\hat{\gamma}_{full} - \gamma_0\right)$ is another linear combination of M and N, namely

$$D = (\delta + W) = \delta + K\left(N - J_{10}J_{00}^{-1}M\right)$$

Therefore, the limiting distribution of all the submodel estimators *jointly* with

D_n can be written as the appropriate linear combination of $(M', N')'$, so the joint distribution is multivariate normal. Now we can apply the continuous mapping theorem as desired, to find:

$$\sqrt{n}\left(\hat{\mu} - \mu_{true}\right) = \sum_S \left[c(S|D_n)\sqrt{n}\left(\hat{\mu}_S - \mu_{true}\right)\right] \xrightarrow{d} \sum_S c(S|D_n)\Lambda_S$$

where Λ_S is the limit distribution of $\sqrt{n}\left(\hat{\mu}_S - \mu_{true}\right)$ defined above. Let

$$\Lambda \equiv \sum_S c(S|D_n)\Lambda_S$$

We want to express Λ in a more convenient form using the fact that

$$\Lambda_S = \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \omega' \left(\delta - K^{1/2} H_S K^{-1/2} D\right)$$

as shown in Lemma 3.3. Substituting,

$$\begin{aligned}
\Lambda &= \sum_S c(S|D)\left[\nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \omega' \left(\delta - K^{1/2} H_S K^{-1/2} D\right)\right] \\
&= \left[\sum_S c(S|D)\right] \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \sum_S \left[c(S|D)\omega' \left(\delta - K^{1/2} H_S K^{-1/2} D\right)\right] \\
&= \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \left[\sum_S c(S|D)\right] \omega'\delta - \sum_S c(S|D)\omega' K^{1/2} H_S K^{-1/2} D \\
&= \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \omega'\delta - \sum_S c(S|D)\omega' K^{1/2} H_S K^{-1/2} D \\
&= \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \omega'\delta - \omega' K^{1/2} \left[\sum_S c(S|D) H_S\right] K^{-1/2} D \\
&= \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \omega'\delta - \omega' \left(K^{-1/2} \left[\sum_S c(S|D) H_S\right] K^{1/2}\right)' D \\
&= \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \omega'\delta - \omega' G(D)' D \\
&= \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \omega'\delta - \omega'\hat{\delta}(D) \\
&= \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \omega' \left\{\delta - \hat{\delta}(D)\right\}
\end{aligned}$$

where we have used the following facts:

1. Only H_S depends on S.
2. The weights sum to one.
3. As scalars, the weights commute and are (trivially) symmetric.
4. H_S, $K^{1/2}$, and $K^{-1/2}$ are symmetric.

along with the definitions of $G(\xrightarrow{d} ot)$ and $\hat{\delta}(D)$. Now we have shown that

$$\Lambda = \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \omega' \left[\delta - \hat{\delta}(D)\right]$$

Notice that, since $\hat{\delta}(D)$ depends only on $D = \delta + W$, and M is independent of W, it follows that the two terms in this expression are likewise independent. The first follows a normal distribution but the second is, in general, non-normal.

Now, since M and $D = \delta + W$ are independent, it follows that the distribution of $M|D$ is the same as that of M. Thus,

$$\Lambda|(D = d) \sim \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \omega' \left[\delta - \hat{\delta}(d)\right]$$

which is a normal distribution, since $\hat{\delta}(d)$ is a constant taking into account the conditioning. The mean and variance are as follows:

$$\begin{aligned} E\left[\Lambda|D = d\right] &= E\left[\nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M\right] + \omega' \left[\delta - \hat{\delta}(d)\right] \\ &= \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} E\left[M\right] + \omega' \left[\delta - \hat{\delta}(d)\right] \\ &= \omega' \left[\delta - \hat{\delta}(d)\right] \end{aligned}$$

$$
\begin{aligned}
Var\left[\Lambda|D=d\right] &= Var\left[\nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}M\right] \\
&= \nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}Var[M]J_{00}^{-1}\nabla_\theta\mu(\theta_0,\gamma_0) \\
&= \nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}J_{00}J_{00}^{-1}\nabla_\theta\mu(\theta_0,\gamma_0) \\
&= \nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}\nabla_\theta\mu(\theta_0,\gamma_0) \\
&\equiv \tau_0^2
\end{aligned}
$$

since $\omega'(\delta - \hat{\delta}(d))$ is a constant. Note that τ_0^2 is the *minimal variance* of the estimators under consideration. Although the *unconditional distribution* of Λ is non-normal, we can still calculate its mean and variance using our decomposition into two independent terms and the linearity of expectation:

$$
\begin{aligned}
E\left[\Lambda\right] &= E\left[\nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}M\right] + E\left[\omega'\left\{\delta - \hat{\delta}(d)\right\}\right] \\
&= \omega'\delta - \omega'E\left[\hat{\delta}(d)\right] \\
&= \omega'\left\{\delta - E\left[\hat{\delta}(d)\right]\right\}
\end{aligned}
$$

$$
\begin{aligned}
Var\left[\Lambda\right] &= Var\left[\nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}M\right] + Var\left[\omega'\left\{\delta - \hat{\delta}(d)\right\}\right] \\
&= \tau_0^2 + \omega' Var\left[\hat{\delta}(D)\right]\omega
\end{aligned}
$$

Now, Λ is the limit distribution of $\sqrt{n}(\hat{\mu} - \mu_{true})$ where $\hat{\mu}$ is the compromise estimator, thus if asymptotically unbiased, it should be centered around zero.

Accordingly we find the MSE of Λ as follows:

$$\begin{aligned}
MSE(\Lambda) &= E\left[(\Lambda-0)^2\right] = E\left[\Lambda^2\right] \\
&= E\left[\left(\nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}M+\omega'\left\{\delta-\hat{\delta}(D)\right\}\right)^2\right] \\
&= E\left[\left\{\nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}M\right\}^2\right]+E\left[\left(\omega'\left\{\delta-\hat{\delta}(D)\right\}\right)^2\right] \\
&\qquad +2E\left[\nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}M\omega'\left\{\delta-\hat{\delta}(D)\right\}\right] \\
&= E\left[\nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}MM'J_{00}^{-1}\nabla_\theta\mu(\theta_0,\gamma_0)\right] \\
&\qquad +E\left[\omega'\{\delta-\hat{\delta}(D)\}\{\delta-\hat{\delta}(D)\}'\omega\right] \\
&\qquad +2E\left[\nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}M\omega'\left\{\delta-\hat{\delta}(D)\right\}\right] \\
&= \nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}E\left[MM'\right]J_{00}^{-1}\nabla_\theta\mu(\theta_0,\gamma_0) \\
&\qquad +\omega'E\left[\{\delta-\hat{\delta}(D)\}\{\delta-\hat{\delta}(D)\}'\right]\omega \\
&\qquad +2\nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}E\left[M\right]E\left[\omega'\left\{\delta-\hat{\delta}(D)\right\}\right] \\
&= \nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}J_{00}J_{00}^{-1}\nabla_\theta\mu(\theta_0,\gamma_0) \\
&\qquad +\omega'E\left[\{\delta-\hat{\delta}(D)\}\{\delta-\hat{\delta}(D)\}'\right]\omega \\
&= \tau_0^2+\omega'E\left[\{\delta-\hat{\delta}(D)\}\{\delta-\hat{\delta}(D)\}'\right]\omega \\
&= \tau_0^2+R(\delta)
\end{aligned}$$

Where we have used the fact that $E[M]=0$ and hence $Var[M]=E[MM']$ and the independence of M and D and hence of any measurable functions thereof. □

IMPORTANT: For convenience, define $\Lambda_0=\nabla_\theta\mu(\theta_0,\gamma_0)'J_{00}^{-1}M$. The *key point* here is that the distribution of

$$\Lambda=\sum_S c(S|D)\Lambda_S=\Lambda_0+\omega'\{\delta-\hat{\delta}(D)\}$$

is often **dramatically non-normal**. To find the density of Λ, first condition on D using the result from above:

$$\begin{aligned} \Lambda|(D=x) &= \Lambda_0 + \omega'\{\delta - \hat{\delta}(x)\} \\ &\sim \mathcal{N}(0, \tau_0^2) + \omega'\{\delta - \hat{\delta}(x)\} \end{aligned}$$

Now, let $h(z)$ denote the density of Λ. We can calculate h by integrating D out of the joint density of (Λ, D). Let $f(x)$ denote the density of D. We have

$$\begin{cases} h(z|D=x) \sim \mathcal{N}\left(\omega'\{\delta - \hat{\delta}(x)\}, \tau_0^2\right) \\ D = \delta + W \sim \mathcal{N}_q(\delta, K) \end{cases}$$

Now factor the joint density according to $h(z|D=x)f(x)$ and integrate out D as follows:

$$h(z) = \int h(z|D=x)f(x)\, dx$$

We can then substitute the two normal distributions and then either numerically integrate or simulate. *Notice*, however, that **the result depends on the unknown constant** δ.

Using the Full Model Variance One approach to constructing a confidence interval that takes account of model selection uncertainty is to essentially use the variance of the full model. Define

$$\begin{aligned} \tau_{full}^2 &= \text{AVAR}(\hat{\mu}_{full}) = Var[\Lambda_{full}] \\ &= \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} \nabla_\theta \mu(\theta_0, \gamma_0) + \omega' K^{1/2} H_{full} K^{1/2} \omega \\ &= \tau_0^2 + \omega' K^{1/2} H_{full} K^{1/2} \omega \\ &= \tau_0^2 + \omega' K^{1/2} \left\{ K^{-1/2} (\pi_{full}' K_{full} \pi_{full}) K^{-1/2} \right\} K^{1/2} \omega \\ &= \tau_0^2 + \omega' K^{1/2} K^{-1/2} K K^{-1/2} K^{1/2} \omega \\ &= \tau_0^2 + \omega' K \omega \end{aligned}$$

And accordingly $\tau_{full} = (\tau_0^2 + \omega' K \omega)^{1/2}$. Now let $\hat{\omega}$ be a consistent estimator of ω and $\widehat{\kappa}$ be a consistent estimator of τ_{full}. Define

$$T_n = \left[\sqrt{n}(\hat{\mu} - \mu_{true}) - \widehat{\omega}' \left\{ D_n - \sum_{S \in \mathcal{A}} c(S|D_n) G_S D_n \right\} \right] \Big/ \widehat{\kappa}$$

From above, we know that the following converges jointly in distribution:

$$\begin{bmatrix} \sqrt{n}(\hat{\mu} - \mu_{true}) \\ D_n \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Lambda_0 + \omega' \left\{ \delta - \sum_{S \in \mathcal{A}} c(S|D) G_S D \right\} \\ D \end{bmatrix}$$

Thus

$$\begin{aligned} T_n &\xrightarrow{d} \left[\Lambda_0 + \omega' \left\{ \delta - \sum_{S \in \mathcal{A}} c(S|D) G_S D \right\} - \omega' \left\{ D - \sum_{S \in \mathcal{A}} c(S|D) G_S D \right\} \right] \Big/ \tau_{full} \\ &= (\tau_0^2 + \omega' K \omega)^{-1/2} \left[\Lambda_0 + \omega' (\delta - D) \right] \end{aligned}$$

We know from above that

$$\begin{bmatrix} M \\ W \end{bmatrix} \sim \mathcal{N}_{p+q} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} J_{00} & 0 \\ 0 & K \end{bmatrix} \right)$$

where $K = J^{11}$, so

$$\begin{aligned} \Lambda_0 &\equiv \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M \\ &\sim \mathcal{N}\left(0, \tau_0^2\right) \end{aligned}$$

independently of

$$\begin{aligned} \omega' (\delta - D) &\equiv -\omega' W \\ &\sim \mathcal{N}(0, \omega' K \omega) \end{aligned}$$

Therefore

$$\begin{aligned} T_n &\xrightarrow{d} (\tau_0^2 + \omega' K \omega)^{-1/2} \left[\Lambda_0 + \omega' \left(\delta - D\right)\right] \\ &= (\tau_0^2 + \omega' K \omega)^{-1/2} \times \mathcal{N}\left(0, \tau_0^2 + \omega' K \omega\right) \\ &\sim \mathcal{N}\left(0, 1\right) \end{aligned}$$

which is a **standard normal**! We can use this result to create a approximate confidence interval for $\hat{\mu}$ as follows. For large n,

$$T_n = \widehat{\kappa}^{-1} \left[\sqrt{n}(\hat{\mu} - \mu_{true}) - \omega' \left\{ D_n - \sum_{S \in \mathcal{A}} c(S|D_n) G_S D_n \right\} \right] \approx \mathcal{N}(0,1)$$

To create a $(1-\alpha) \times 100\%$ interval, define $z_{\alpha/2}$ as the appropriate quantile of a standard normal random variable so that

$$P\left[-z_{\alpha/2} \leq T_n \leq z_{\alpha/2}\right] \approx 1 - \alpha$$

Then,

$$
\begin{aligned}
1-\alpha &\approx P\left[-\frac{z_{\alpha/2}\widehat{\kappa}}{\sqrt{n}} \leq (\hat{\mu}-\mu_{true}) - \frac{\widehat{\omega}'}{\sqrt{n}}\left\{D_n - \sum_{S\in\mathcal{A}} c(S|D_n)G_S D_n\right\} \leq \frac{z_{\alpha/2}\widehat{\kappa}}{\sqrt{n}}\right] \\
&= P\left[\frac{z_{\alpha/2}\widehat{\kappa}}{\sqrt{n}} \geq (\mu_{true}-\hat{\mu}) + \frac{\widehat{\omega}'}{\sqrt{n}}\left\{D_n - \sum_{S\in\mathcal{A}} c(S|D_n)G_S D_n\right\} \geq -\frac{z_{\alpha/2}\widehat{\kappa}}{\sqrt{n}}\right] \\
&= P\left[-\frac{z_{\alpha/2}\widehat{\kappa}}{\sqrt{n}} \leq (\mu_{true}-\hat{\mu}) + \frac{\widehat{\omega}'}{\sqrt{n}}\left\{D_n - \sum_{S\in\mathcal{A}} c(S|D_n)G_S D_n\right\} \leq \frac{z_{\alpha/2}\widehat{\kappa}}{\sqrt{n}}\right] \\
&= P\left[\hat{\mu}-\frac{z_{\alpha/2}\widehat{\kappa}}{\sqrt{n}} \leq \mu_{true} + \frac{\widehat{\omega}'}{\sqrt{n}}\left\{D_n - \sum_{S\in\mathcal{A}} c(S|D_n)G_S D_n\right\} \leq \hat{\mu}+\frac{z_{\alpha/2}\widehat{\kappa}}{\sqrt{n}}\right] \\
&= P\left[\hat{\mu}-\frac{z_{\alpha/2}\widehat{\kappa}}{\sqrt{n}} \leq \mu_{true} + \Delta_n \leq \hat{\mu}+\frac{z_{\alpha/2}\widehat{\kappa}}{\sqrt{n}}\right] \\
&= P\left[\left(\hat{\mu}-\frac{z_{\alpha/2}\widehat{\kappa}}{\sqrt{n}}\right) - \Delta_n \leq \mu_{true} \leq \left(\hat{\mu}+\frac{z_{\alpha/2}\widehat{\kappa}}{\sqrt{n}}\right) - \Delta_n\right]
\end{aligned}
$$

where

$$\Delta_n \equiv \frac{\widehat{\omega}'}{\sqrt{n}}\left[D_n - \sum_{S\in\mathcal{A}} c(S|D_n)G_S D_n\right]$$

According to Claeskens and Hjort: "this method is first-order equivalent to using the full model for confidence interval construction, with a modification for location."

Correcting Confidence Intervals: Simulation Another possibility is to simulate from the limiting distribution of Λ *for a range fixed value of* δ, using consistent estimates of all other unknown quantities. This procedure can then be repeated for a variety of choices of δ. To make this clearer, first rewrite Λ

using an expression from the proof Theorem 4.1

$$\begin{aligned}
\Lambda &= \nabla_\theta \mu(\theta_0, \gamma_0)' J_{00}^{-1} M + \omega' \delta - \omega' K^{1/2} \left[\sum_S c(S|D) H_S \right] K^{-1/2} D \\
&= \Lambda_0 + \omega' \left[\delta - K^{1/2} \sum_S c(S|D) H_S K^{-1/2} D \right] \\
&= \Lambda_0 + \omega' \left[\delta - \sum_S c(S|D) K^{1/2} H_S K^{-1/2} D \right] \\
&= \Lambda_0 + \omega' \left[\delta - \sum_S c(S|D) G_S D \right]
\end{aligned}$$

where we have defined $G_S = K^{1/2} H_S K^{-1/2}$. We know from above that $\Lambda_0 \sim \mathcal{N}(0, \tau_0^2)$ independent of $D \sim \mathcal{N}_q(\delta, K)$. The simulation procedure is as follows:

1. Calculate consistent estimates of G_S, τ_0^2, ω and K using the estimation results and fix a value of δ.

2. Using the result of step one, generate:

   (a) $\Lambda_{0,j} \sim \mathcal{N}(0, \hat{\tau}_0^2)$ independently of

   (b) $D_j \sim \mathcal{N}_q(\delta, \hat{K})$

3. Calculate the weights c using D_j and set

$$\Lambda_j = \Lambda_{0,j} + \hat{\omega}' \left[\delta - \sum_S c(S|D_j) \hat{G}_S D_j \right]$$

4. Repeat steps 1 and 2 for $j = 1, 2, \ldots, B$

5. Using the samples $\{\Lambda_1, \Lambda_2, \ldots, \Lambda_B\}$ generated in steps 3 and 4, calculate quantiles $a(\delta)$ and $b(\delta)$ that satisfy:

$$P\left[a(\delta) \leq \Lambda(\delta) \leq b(\delta)\right] = 0.95$$

6. Repeat steps 2–5 for varying choices of δ.

Now, suppose we know the values $[a(\delta), b(\delta)]$. Since Λ is the limit distribution of $\sqrt{n}(\hat{\mu} - \mu_{true})$, it follows that

$$P\left[a(\delta) \leq \sqrt{n}(\hat{\mu} - \mu_{true}) \leq b(\delta)\right] \to P\left[a(\delta) \leq \Lambda(\delta) \leq b(\delta)\right] = 0.95$$

Thus, $[\hat{\mu} - b(\delta)/\sqrt{n}, \hat{\mu} - a(\delta)/\sqrt{n}]$ covers μ_{true} with probability 0.95 asymptotically.

But We Don't Know δ! A naive approach would be to substitute our estimate $D_n = \hat{\delta}_{full} = \sqrt{n}(\hat{\gamma}_{full} - \gamma_0)$, carrying about the above simulations at this value and creating an interval based on $\hat{a} = a(D_n)$ and $\hat{b} = b(D_n)$. This is simple, but may not always work well. Let $p_n(\delta)$ be the coverage probability for this procedure. Its limit is

$$p(\delta) = P\left[a(D) \leq \Lambda(\delta) \leq b(D)\right]$$

which can be simulated by the method described above. It turns out that this method sometimes gives coverage that is *far too low*.

A Better Procedure: Instead of simply substituting D_n for δ in the simulation described above, we could first construct a confidence region for δ and use this region to create an interval for $\hat{\mu}$. Since

$$D_n \xrightarrow{d} D = \delta + W \sim \mathcal{N}_q\left(\delta, K\right)$$

where $K = J^{11}$, we have

$$\left(D_n - \delta\right)' \widehat{K}^{-1} \left(D_n - \delta\right) \xrightarrow{d} \chi^2_q$$

Now, define

$$\rho_n(D_n, \delta) = \left[\left(D_n - \delta\right)' \widehat{K}^{-1} \left(D_n - \delta\right)\right]^{1/2}$$

and the event

$$A_n(c) = \{\rho_n(D_n, \delta) \leq c\}$$

Now, since $\rho_n(D_n, \delta)^2 \approx \chi^2_q$ we have

$$P\{A_n(c)\} = P\{\rho_n(D_n, \delta) \leq c\} = P\left\{\rho_n(D_n, \delta)^2 \leq c^2\right\} \approx P\left\{\chi^2_q \leq c^2\right\}$$

where we have used the fact that x^2 is strictly increasing for $x \geq 0$ and that $\rho_n \geq 0$. Now define $z = (\chi^2_{q,0.95})^{1/2}$ and $A_n = A_n(z)$, so that $P\{A_n\} \approx 0.95$. In the simulations described above in which we assumed that δ was known, we defined $a(\delta)$ and $b(\delta)$ so that

$$P[a(\delta) \leq \Lambda(\delta) \leq b(\delta)] = 0.95$$

Now, let

$$\begin{aligned} \widehat{a}_0(D_n) &= \min\{a(\delta)\colon \rho_n(D_n, \delta) \leq z\} \\ \widehat{b}_0(D_n) &= \max\{b(\delta)\colon \rho_n(D_n, \delta) \leq z\} \end{aligned}$$

The claim is that the limit coverage level of

$$\mathrm{CI}^*_n = \left[\widehat{\mu} - \frac{\widehat{b}_0(D_n)}{\sqrt{n}}, \quad \widehat{\mu} - \frac{\widehat{a}_0(D_n)}{\sqrt{n}}\right]$$

is always *above* 0.90, resulting in a conservative procedure. To see why this is the case, we return to the limit experiment, in which we have joint convergence of all the necessary random variables, as described above. This implies that the coverage probability $r_n(\delta)$ to which $\{\mu_{true} \in \mathrm{CI}^*_n\}$ converges is given by

$$r(\delta) = P\{a_0(D) \leq \Lambda(\delta) \leq b_0(D)\}$$

where $\rho(D,\delta)^2 = (D-\delta)' K^{-1} (D-\delta)$ and

$$\begin{aligned} a_0(D) &= \min\{a(\delta)\colon \rho(D,\delta) \leq z\} \\ b_0(D) &= \min\{b(\delta)\colon \rho(D,\delta) \leq z\} \end{aligned}$$

How does this work? The interval CI_n^* is based on

$$\begin{aligned} 0.9 &\leq P\left[\widehat{\mu} - \frac{\widehat{b}_0(D_n)}{\sqrt{n}} \leq \mu_{true} \leq \widehat{\mu} - \frac{\widehat{a}_0(D_n)}{\sqrt{n}}\right] \\ &= P\left[-\frac{\widehat{b}_0(D_n)}{\sqrt{n}} \leq \mu_{true} - \widehat{\mu} \leq -\frac{\widehat{a}_0(D_n)}{\sqrt{n}}\right] \\ &= P\left[-\widehat{b}_0(D_n) \leq \sqrt{n}\,(\mu_{true} - \widehat{\mu}) \leq -\widehat{a}_0(D_n)\right] \\ &= P\left[\widehat{b}_0(D_n) \geq \sqrt{n}\,(\widehat{\mu} - \mu_{true}) \geq \widehat{a}_0(D_n)\right] \\ &= P\left[\widehat{a}_0(D_n) \leq \sqrt{n}\,(\widehat{\mu} - \mu_{true}) \leq \widehat{b}_0(D_n)\right] \end{aligned}$$

Now, we know that

$$\begin{bmatrix} \sqrt{n}(\hat{\mu} - \mu_{true}) \\ D_n \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Lambda_0 + \omega' \left\{\delta - \sum_{S\in\mathcal{A}} c(S|D) G_S D\right\} \\ D \end{bmatrix}$$

so, by the Continuous Mapping Theorem,

$$\begin{bmatrix} \sqrt{n}(\hat{\mu} - \mu_{true}) \\ \rho_n(D_n,\delta) \\ \widehat{a}_0(D_n) \\ \widehat{b}_0(D_n) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \Lambda(\delta) \\ \rho(D,\delta) \\ a_0(D) \\ b_0(D) \end{bmatrix}$$

and thus,

$$P\left[\widehat{a}_0(D_n) \leq \sqrt{n}\,(\widehat{\mu} - \mu_{true}) \leq \widehat{b}_0(D_n)\right] \to P\left\{a_0(D) \leq \Lambda(\delta) \leq b_0(D)\right\} = r(\delta)$$

Now, let $A = \{\rho(D, \delta) \leq z\}$ where, as before, $z = (\chi^2_{q,0.95})^{1/2}$ implying that $P\{A\} = 0.95$. Then,

$$\begin{aligned} 0.95 &= P\{a(\delta) \leq \Lambda(\delta) \leq b(\delta)\} \\ &= P[\{a(\delta) \leq \Lambda(\delta) \leq b(\delta)\} \cap A] + P[\{a(\delta) \leq \Lambda(\delta) \leq b(\delta)\} \cap A^c] \end{aligned}$$

Now, since

$$\begin{aligned} a_0(D) &= \min\{a(\delta) \colon \rho(D, \delta) \leq z\} \\ b_0(D) &= \min\{b(\delta) \colon \rho(D, \delta) \leq z\} \end{aligned}$$

we have

$$\{a(\delta) \leq \Lambda(\delta) \leq b(\delta)\} \cap A \Rightarrow \{a_0(D) \leq \Lambda(\delta) \leq b_0(D)\}$$

and hence

$$P[\{a(\delta) \leq \Lambda(\delta) \leq b(\delta)\} \cap A] \leq P\{a_0(D) \leq \Lambda(\delta) \leq b_0(D)\}$$

Further, since

$$\{a(\delta) \leq \Lambda(\delta) \leq b(\delta)\} \cap A^c \Rightarrow A^c$$

we have

$$P[\{a(\delta) \leq \Lambda(\delta) \leq b(\delta)\} \cap A^c] \leq P(A^c)$$

Combining:

$$\begin{aligned} 0.95 &= P[\{a(\delta) \leq \Lambda(\delta) \leq b(\delta)\} \cap A] + P[\{a(\delta) \leq \Lambda(\delta) \leq b(\delta)\} \cap A^c] \\ &\leq P\{a_0(D) \leq \Lambda(\delta) \leq b_0(D)\} + P(A^c) \\ &= P\{a_0(D) \leq \Lambda(\delta) \leq b_0(D)\} + 0.05 \end{aligned}$$

since A is defined with reference to a 95% confidence interval. Subtracting,

$$P\left\{a_0(D) \leq \Lambda(\delta) \leq b_0(D)\right\} \geq 0.90$$

as claimed.

Here's the intuition for what just happened. Λ is a random variable whose distribution depends on the unknown constant δ. The constants $a(\delta)$ and $b(\delta)$ are quantiles of the distribution of Λ such that

$$P\left[a(\delta) \leq \Lambda(\delta) \leq b(\delta)\right] = 0.95$$

Since Λ depends on δ, so do its quantiles: different values of δ would result in different intervals.

This is the procedure. First we use $\rho(D, \delta) \leq z$ to get a confidence interval for δ. Then we plug each point in this interval for δ into $\Lambda(\delta)$ and calculate the corresponding bounds $a(\delta)$ and $b(\delta)$. For each value of δ such that $\rho(D, \delta) \leq z$ we get a *different* confidence interval for Λ. The lower bound of all these intervals is $a_0(D)$ while the upper bound is $b_0(D)$. The point here is to assess the coverage of the resulting interval.

The confusion here comes from bad notation: sometimes δ is being treated as fixed, other times as variable. Need to come up with some clearer notation...