Political Methodology III: Model Based Inference

Justin Grimmer

Professor
Department of Political Science
Stanford University

April 16th, 2019

Most Important Problems

Effects of presidents going public

- $Y_i = 1$ if respondent identifies MIP as topic of president's speech
- $\boldsymbol{X}_i = (1, \text{Treatment}, \text{Republican})$

Review: Probit

$$\underbrace{Y_i}_{\text{Random}} \sim \text{Bernoulli}(\pi_i)$$

Review: Probit

$$\underbrace{Y_i}_{\text{Random}} \sim \text{Bernoulli}(\pi_i)$$

$$\pi_i = \Phi(\boldsymbol{X}_i^{'} \underbrace{\boldsymbol{\beta}}_{\text{Infer (Fixed)}})$$

Review: Probit

$$
\begin{aligned}
\underbrace{Y_i}_{\text{Random}} & \sim \text{Bernoulli}(\pi_i) \\
\pi_i & = \Phi(\boldsymbol{X}_i^{'} \underbrace{\boldsymbol{\beta}}_{\text{Infer (Fixed)}})
\end{aligned}
$$

$$
L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) = f(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X})
$$

Review: Probit

$$
\begin{aligned}
\underbrace{Y_i}_{\text{Random}} & \sim \text{Bernoulli}(\pi_i) \\
\pi_i & = \Phi(\boldsymbol{X}_i^{'} \underbrace{\boldsymbol{\beta}}_{\text{Infer (Fixed)}})
\end{aligned}
$$

$$
\begin{aligned}
L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) & = f(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X}) \\
& = \prod_{i=1}^{N} \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta})^{Y_i}(1 - \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}))^{1-Y_i}
\end{aligned}
$$

Review: Probit

$$
\begin{aligned}
\underbrace{Y_i}_{\text{Random}} & \sim \text{Bernoulli}(\pi_i) \\
\pi_i & = \Phi(\boldsymbol{X}_i^{'} \underbrace{\boldsymbol{\beta}}_{\text{Infer (Fixed)}})
\end{aligned}
$$

$$
\begin{aligned}
L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) & = f(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X}) \\
& = \prod_{i=1}^{N} \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta})^{Y_i}(1 - \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}))^{1-Y_i} \\
\log L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) & = \sum_{i=1}^{N} \left(Y_i \log \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}) + (1 - Y_i) \log(1 - \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}))\right)
\end{aligned}
$$

Review: Probit

$$\underbrace{Y_i}_{\text{Random}} \sim \text{Bernoulli}(\pi_i)$$

$$\pi_i = \Phi(\boldsymbol{X}_i^{'} \underbrace{\boldsymbol{\beta}}_{\text{Infer (Fixed)}})$$

$$\begin{aligned}
L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) &= f(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{X}) \\
&= \prod_{i=1}^{N} \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta})^{Y_i}(1 - \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}))^{1-Y_i} \\
\log L(\boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Y}) &= \sum_{i=1}^{N} \left(Y_i \log \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}) + (1 - Y_i) \log(1 - \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}))\right) \\
\boldsymbol{\beta}^* &= \boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y}) = \text{Maximum Likelihood Estimator}
\end{aligned}$$

Review: Probit

$$\underbrace{Y_i}_{\text{Random}} \sim \text{Bernoulli}(\pi_i)$$

$$\pi_i = \Phi(\boldsymbol{X}_i^{'} \underbrace{\boldsymbol{\beta}}_{\text{Infer (Fixed)}})$$

$$\begin{aligned}
L(\boldsymbol{\beta}|\boldsymbol{X},\boldsymbol{Y}) &= f(\boldsymbol{Y}|\boldsymbol{\beta},\boldsymbol{X}) \\
&= \prod_{i=1}^{N} \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta})^{Y_i}(1-\Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}))^{1-Y_i} \\
\log L(\boldsymbol{\beta}|\boldsymbol{X},\boldsymbol{Y}) &= \sum_{i=1}^{N}\Big(Y_i \log \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}) + (1-Y_i)\log(1-\Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}))\Big) \\
\boldsymbol{\beta}^* &= \boldsymbol{\beta}^*(\boldsymbol{X},\boldsymbol{Y}) = \text{Maximum Likelihood Estimator} \\
\widehat{\boldsymbol{\beta}} &= \boldsymbol{\beta}^*(\boldsymbol{X},\boldsymbol{y})
\end{aligned}$$

Review: Probit

$$\boldsymbol{\beta}^* \quad \rightarrow^D \quad \text{Multivariate Normal}\left(\boldsymbol{\beta}, I_N(\widehat{\boldsymbol{\beta}})^{-1}\right)$$

Review: Probit

$$\boldsymbol{\beta}^* \quad \rightarrow^D \quad \text{Multivariate Normal}\left(\boldsymbol{\beta}, I_N(\widehat{\boldsymbol{\beta}})^{-1}\right)$$

$$I_N(\widehat{\boldsymbol{\beta}})^{-1} \quad = \quad \begin{pmatrix} \text{var}(\beta_0) & \text{cov}(\beta_0, \beta_1) & \ldots & \text{cov}(\beta_0, \beta_K) \\ \text{cov}(\beta_1, \beta_0) & \text{var}(\beta_1) & \ldots & \text{cov}(\beta_1, \beta_K) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\beta_K, \beta_0) & \text{cov}(\beta_K, \beta_1) & \ldots & \text{var}(\beta_K) \end{pmatrix}$$

Review Probit

Review Probit

For a probit regression coefficient:

Review Probit

For a probit regression coefficient:

$$\mathsf{CI}_{95} = [\widehat{\beta}_k - 1.96 \times \sqrt{\mathsf{var}(\beta_k)}, \widehat{\beta}_k + 1.96 \times \sqrt{\mathsf{var}(\beta_k)}]$$

Review Probit

For a probit regression coefficient:

$$\begin{aligned}
\mathsf{CI}_{95} &= [\widehat{\beta}_k - 1.96 \times \sqrt{\mathsf{var}(\beta_k)}, \widehat{\beta}_k + 1.96 \times \sqrt{\mathsf{var}(\beta_k)}] \\
\mathsf{CI}_{1-\alpha} &= [\widehat{\beta}_k - z_{\alpha/2} \times \sqrt{\mathsf{var}(\beta_k)}, \widehat{\beta}_k + z_{\alpha/2} \times \sqrt{\mathsf{var}(\beta_k)}]
\end{aligned}$$

Review Probit

For a probit regression coefficient:

$$
\begin{aligned}
\mathsf{CI}_{95} &= [\widehat{\beta}_k - 1.96 \times \sqrt{\mathsf{var}(\beta_k)}, \widehat{\beta}_k + 1.96 \times \sqrt{\mathsf{var}(\beta_k)}] \\
\mathsf{CI}_{1-\alpha} &= [\widehat{\beta}_k - z_{\alpha/2} \times \sqrt{\mathsf{var}(\beta_k)}, \widehat{\beta}_k + z_{\alpha/2} \times \sqrt{\mathsf{var}(\beta_k)}]
\end{aligned}
$$

Recall: If $Z \sim \mathsf{Normal}(0, 1)$, $z_{\alpha/2}$ is the number such that

$$
P(|Z \geq z_{\alpha/2}|) = \alpha
$$

Review Probit

```
>colnames(mips)

[1] "mip" "treat" "gop"


>reg1<- glm(mip~treat + gop, family = binomial(link='probit')
data = as.data.frame(mips))

> reg1$coef
(Intercept) treat gop
-0.259173644 -0.009598938 -0.198126517


> sqrt(diag(vcov(reg1)))
(Intercept) treat gop
0.03297588 0.05268254 0.05115777
```

Review Probit

95-percent confidence interval:

```
> ci_treat<- c(reg1$coef[2] - 1.96sqrt(diag(vcov(reg1)))[2],
reg1$coef[2] + 1.96sqrt(diag(vcov(reg1)))[2])

> ci_treat
```

```
-0.11285672 0.09365884
```

Review Probit

Quantities of interest:

Review Probit

Quantities of interest:

$$E[Y_i|\boldsymbol{X}_i] = \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta})$$

Review Probit

Quantities of interest:

$$\begin{aligned} E[Y_i|\boldsymbol{X}_i] &= \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}) \\ E[Y_i|X_{i1} = 1, \boldsymbol{X}_i] - E[Y_i|\tilde{X}_{i1} = 0, \tilde{\boldsymbol{X}}_i] &= \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}) - \Phi(\tilde{\boldsymbol{X}}_i^{'}\boldsymbol{\beta}) \end{aligned}$$

Review Probit

Quantities of interest:

$$
\begin{aligned}
E[Y_i|\boldsymbol{X}_i] &= \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}) \\
E[Y_i|X_{i1}=1,\boldsymbol{X}_i] - E[Y_i|\tilde{X}_{i1}=0,\tilde{\boldsymbol{X}}_i] &= \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}) - \Phi(\tilde{\boldsymbol{X}}_i^{'}\boldsymbol{\beta}) \\
\frac{\partial E[Y_i|X_{i1}=1,\boldsymbol{X}_i]}{\partial X_{i1}} &= \phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta})\beta_1
\end{aligned}
$$

```
> X_synth<- c(1, 0, 1)

> y.tilde<- reg1$coef%*%X_synth

> y.prob<- pnorm(y.tilde)

> y.prob

[1,] 0.3237277
```

Uncertainty for Maximum Likelihood Estimation

Uncertainty for Maximum Likelihood Estimation

1) Coefficients: Asymptotic result $\rightsquigarrow$ YMMV in small samples

Uncertainty for Maximum Likelihood Estimation

1) Coefficients: Asymptotic result $\rightsquigarrow$ YMMV in small samples
2) And even if asymptotic distribution holds:

Uncertainty for Maximum Likelihood Estimation

1) Coefficients: Asymptotic result $\rightsquigarrow$ YMMV in small samples
2) And even if asymptotic distribution holds:
    1) Difficult substantive interpretation of coefficients

Uncertainty for Maximum Likelihood Estimation

1) Coefficients: Asymptotic result $\rightsquigarrow$ YMMV in small samples
2) And even if asymptotic distribution holds:
    1) Difficult substantive interpretation of coefficients
    2) Uncertainty for functions of coefficients

Inference Three Ways

How do obtain uncertainty estimates for Quantities of Interest?

1) Bootstrap $\rightsquigarrow$ (no asymptotics, simulation)
2) Delta Method $\rightsquigarrow$ (asymptotic normality, analytic)
3) Simulation from Multivariate Normal $\rightsquigarrow$ (asymptotic normality, simulation)

The Bootstrap (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ for N observations

For each of M simulations $(m = 1, 2, \ldots, M)$:

The Bootstrap (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ for N observations

For each of M simulations $(m = 1, 2, \ldots, M)$:

1) Sample with replacement N times:

The Bootstrap (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ for N observations

For each of M simulations $(m = 1, 2, \ldots, M)$:

1) Sample with replacement N times:

$$(\tilde{Y}_1, \tilde{\boldsymbol{X}}_1), (\tilde{Y}_2, \tilde{\boldsymbol{X}}_2) \quad \ldots \quad (\tilde{Y}_N, \tilde{\boldsymbol{X}}_N) = (\tilde{Y}, \tilde{\boldsymbol{X}})$$

The Bootstrap (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ for N observations

For each of M simulations $(m = 1, 2, \ldots, M)$:

1) Sample with replacement N times:

$$(\tilde{Y}_1, \tilde{\boldsymbol{X}}_1), (\tilde{Y}_2, \tilde{\boldsymbol{X}}_2) \quad \ldots \quad (\tilde{Y}_N, \tilde{\boldsymbol{X}}_N) = (\tilde{Y}, \tilde{\boldsymbol{X}})$$

2) Calculate $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ and/or any other function of the estimator (expected value, first difference) on bootstrapped sample

The Bootstrap (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ for N observations

For each of M simulations $(m = 1, 2, \ldots, M)$:

1) Sample with replacement N times:

$$(\tilde{Y}_1, \tilde{\boldsymbol{X}}_1), (\tilde{Y}_2, \tilde{\boldsymbol{X}}_2) \quad \ldots \quad (\tilde{Y}_N, \tilde{\boldsymbol{X}}_N) = (\tilde{Y}, \tilde{\boldsymbol{X}})$$

2) Calculate $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ and/or any other function of the estimator (expected value, first difference) on bootstrapped sample

$$\boldsymbol{\beta}^m \quad = \quad \boldsymbol{\beta}^*(\tilde{Y}, \tilde{\boldsymbol{X}})$$

The Bootstrap (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ for N observations

For each of M simulations $(m = 1, 2, \ldots, M)$:

1) Sample with replacement N times:

$$(\tilde{Y}_1, \tilde{\boldsymbol{X}}_1), (\tilde{Y}_2, \tilde{\boldsymbol{X}}_2) \quad \ldots \quad (\tilde{Y}_N, \tilde{\boldsymbol{X}}_N) = (\tilde{Y}, \tilde{\boldsymbol{X}})$$

2) Calculate $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ and/or any other function of the estimator (expected value, first difference) on bootstrapped sample

$$\begin{aligned} \boldsymbol{\beta}^m &= \boldsymbol{\beta}^*(\tilde{Y}, \tilde{\boldsymbol{X}}) \\ h(\boldsymbol{\beta})^m &= h(\boldsymbol{\beta}^*(\tilde{Y}, \tilde{\boldsymbol{X}})) \end{aligned}$$

The Bootstrap (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ for N observations

For each of M simulations $(m = 1, 2, \ldots, M)$:

1) Sample with replacement N times:

$$(\tilde{Y}_1, \tilde{\boldsymbol{X}}_1), (\tilde{Y}_2, \tilde{\boldsymbol{X}}_2) \quad \ldots \quad (\tilde{Y}_N, \tilde{\boldsymbol{X}}_N) = (\tilde{Y}, \tilde{\boldsymbol{X}})$$

2) Calculate $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ and/or any other function of the estimator (expected value, first difference) on bootstrapped sample

$$\begin{aligned} \boldsymbol{\beta}^m &= \boldsymbol{\beta}^*(\tilde{Y}, \tilde{\boldsymbol{X}}) \\ h(\boldsymbol{\beta})^m &= h(\boldsymbol{\beta}^*(\tilde{Y}, \tilde{\boldsymbol{X}})) \end{aligned}$$

3) Do following:

The Bootstrap (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ for N observations

For each of M simulations $(m = 1, 2, \ldots, M)$:

1) Sample with replacement N times:

$$(\tilde{Y}_1, \tilde{\boldsymbol{X}}_1), (\tilde{Y}_2, \tilde{\boldsymbol{X}}_2) \quad \ldots \quad (\tilde{Y}_N, \tilde{\boldsymbol{X}}_N) = (\tilde{Y}, \tilde{\boldsymbol{X}})$$

2) Calculate $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ and/or any other function of the estimator (expected value, first difference) on bootstrapped sample

$$\begin{aligned} \boldsymbol{\beta}^m &= \boldsymbol{\beta}^*(\tilde{Y}, \tilde{\boldsymbol{X}}) \\ h(\boldsymbol{\beta})^m &= h(\boldsymbol{\beta}^*(\tilde{Y}, \tilde{\boldsymbol{X}})) \end{aligned}$$

3) Do following:

$$E[\boldsymbol{\beta}^*] \approx \sum_{m=1}^{M} \frac{\boldsymbol{\beta}^m}{M}$$

The Bootstrap (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ for N observations

For each of M simulations $(m = 1, 2, \ldots, M)$:

1) Sample with replacement N times:

$$(\tilde{Y}_1, \tilde{\boldsymbol{X}}_1), (\tilde{Y}_2, \tilde{\boldsymbol{X}}_2) \quad \ldots \quad (\tilde{Y}_N, \tilde{\boldsymbol{X}}_N) = (\tilde{Y}, \tilde{\boldsymbol{X}})$$

2) Calculate $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ and/or any other function of the estimator (expected value, first difference) on bootstrapped sample

$$\begin{aligned} \boldsymbol{\beta}^m &= \boldsymbol{\beta}^*(\tilde{Y}, \tilde{\boldsymbol{X}}) \\ h(\boldsymbol{\beta})^m &= h(\boldsymbol{\beta}^*(\tilde{Y}, \tilde{\boldsymbol{X}})) \end{aligned}$$

3) Do following:

$$\text{Var}(\boldsymbol{\beta}^*) \approx \sum_{m=1}^{M} \frac{\left(\boldsymbol{\beta}^m - \left(\sum_{m=1}^{M} \frac{\boldsymbol{\beta}^m}{M}\right)\right)^2}{M}$$

The Bootstrap (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ for N observations

For each of M simulations $(m = 1, 2, \ldots, M)$:

1) Sample with replacement N times:

$$(\tilde{Y}_1, \tilde{\boldsymbol{X}}_1), (\tilde{Y}_2, \tilde{\boldsymbol{X}}_2) \quad \ldots \quad (\tilde{Y}_N, \tilde{\boldsymbol{X}}_N) = (\tilde{Y}, \tilde{\boldsymbol{X}})$$

2) Calculate $\boldsymbol{\beta}^*(\boldsymbol{X}, \boldsymbol{Y})$ and/or any other function of the estimator (expected value, first difference) on bootstrapped sample

$$\begin{aligned} \boldsymbol{\beta}^m &= \boldsymbol{\beta}^*(\tilde{Y}, \tilde{\boldsymbol{X}}) \\ h(\boldsymbol{\beta})^m &= h(\boldsymbol{\beta}^*(\tilde{Y}, \tilde{\boldsymbol{X}})) \end{aligned}$$

3) Do following:

$$E[h(\boldsymbol{\beta}^*)] \approx \sum_{m=1}^{M} \frac{h(\boldsymbol{\beta})^m}{M}$$

Example:

```
store_expected<- c()
for(z in 1:1000){
    subset<- sample(1:nrow(mips), nrow(mips), replace=T)
    use_mips<- mips[subset,]
    temp_reg<- glm(mip~treat + gop,
       family = binomial(link='probit'),
       data = as.data.frame(use_mips))
    store_expected[z]<- pnorm(temp_reg$coef%*%X_synth)
}
> mean(store_expected) ## estimate of expected value
[1] 0.3245165
> sd(store_expected) ## estimate of standard error
[1] 0.0157724
```

The Bootstrap: Justification

Two key ideas:

1) Simulation
2) Approximation of cumulative distribution function F with empirical distribution function $\hat{F}_n$

Simulation

- Sample M draws, $Y_m \sim F$
- For example $Y_m \sim \text{Normal}(\mu, \sigma^2)$

Simulation

- Sample M draws, $Y_m \sim F$
- For example $Y_m \sim \text{Normal}(\mu, \sigma^2)$

$$\frac{1}{M}\sum_{m=1}^{M} Y_m \quad \longrightarrow^{P} \quad E[Y] = \int y f(y) dy = \int y dF(y)$$

Simulation

- Sample M draws, $Y_m \sim F$
- For example $Y_m \sim \text{Normal}(\mu, \sigma^2)$

$$\frac{1}{M}\sum_{m=1}^{M} Y_m \quad \longrightarrow^{P} \quad E[Y] = \int y f(y) dy = \int y dF(y)$$

$$\frac{1}{M}\sum_{m=1}^{M} h(Y_m) \quad \longrightarrow^{P} \quad E[h(Y)] = \int h(y) f(y) dy$$

Simulation

- Sample M draws, $Y_m \sim F$
- For example $Y_m \sim \text{Normal}(\mu, \sigma^2)$

$$
\begin{aligned}
\frac{1}{M}\sum_{m=1}^{M} Y_m \quad & \longrightarrow^{P} \quad E[Y] = \int y f(y) dy = \int y dF(y) \\
\frac{1}{M}\sum_{m=1}^{M} h(Y_m) \quad & \longrightarrow^{P} \quad E[h(Y)] = \int h(y) f(y) dy \\
\frac{1}{M}\sum_{m=1}^{M} \left(Y_m - \sum_{m=1}^{M} \frac{Y_m}{M} \right)^2 \quad & \longrightarrow^{P} \quad \int (y-\mu)^2 f(y) dy
\end{aligned}
$$

Empirical Distribution Function

Empirical Distribution Function

Univariate version: X_i is a scalar

$$\hat{F}_n(x) = \sum_{i=1}^{N} \frac{I(X_i \leq x)}{N}$$

Empirical Distribution Function

Univariate version: X_i is a scalar

$$\hat{F}_n(x) = \sum_{i=1}^{N} \frac{I(X_i \leq x)}{N}$$

Multivariate Version: $\boldsymbol{X}_i$ is a vector

$$\hat{F}_n(\boldsymbol{x}) = \sum_{i=1}^{N} \frac{I(\boldsymbol{X}_i \leq \boldsymbol{x})}{N}$$

Empirical Distribution Function

Univariate version: X_i is a scalar

$$\hat{F}_n(x) = \sum_{i=1}^{N} \frac{I(X_i \leq x)}{N}$$

Multivariate Version: $\boldsymbol{X}_i$ is a vector

$$\hat{F}_n(\boldsymbol{x}) = \sum_{i=1}^{N} \frac{I(\boldsymbol{X}_i \leq \boldsymbol{x})}{N}$$

Empirical density function: every observation has density height (probability) $\frac{1}{N}$.

Approximation of Cumulative Distribution Function F

Expected value depends on F

Approximation of Cumulative Distribution Function F

Expected value depends on F

$$\text{Mean}_F(Y) = E[Y] = \int y dF(y)$$

Approximation of Cumulative Distribution Function F

Expected value depends on F

$$
\begin{aligned}
\mathsf{Mean}_F(Y) &= E[Y] = \int y dF(y) \\
\mathsf{Var}_F(Y) &= E[(Y-\mu)^2] = \int (y-\mu)^2 dF(y)
\end{aligned}
$$

Approximation of Cumulative Distribution Function F

Expected value depends on F

$$\begin{aligned}
\text{Mean}_F(Y) &= E[Y] = \int y dF(y) \\
\text{Var}_F(Y) &= E[(Y-\mu)^2] = \int (y-\mu)^2 dF(y) \\
\text{h}_F(Y) &= E[h(Y)] = \int h(y) dF(y)
\end{aligned}$$

Approximation of Cumulative Distribution Function F

Expected value depends on F

$$
\begin{aligned}
\mathsf{Mean}_F(Y) &= E[Y] = \int y dF(y) \\
\mathsf{Var}_F(Y) &= E[(Y-\mu)^2] = \int (y-\mu)^2 dF(y) \\
\mathsf{h}_F(Y) &= E[h(Y)] = \int h(y) dF(y) \\
\mathsf{Mean}_F(Y) &\approx \mathsf{Mean}_{\hat{F}}(Y)) = \int y d\hat{F}(y)
\end{aligned}
$$

Approximation of Cumulative Distribution Function F

Expected value depends on F

$$\begin{aligned}
\text{Mean}_F(Y) &= E[Y] = \int y dF(y) \\
\text{Var}_F(Y) &= E[(Y-\mu)^2] = \int (y-\mu)^2 dF(y) \\
\text{h}_F(Y) &= E[h(Y)] = \int h(y) dF(y) \\
\text{Mean}_F(Y) &\approx \text{Mean}_{\hat{F}}(Y)) = \int y d\hat{F}(y) \\
\text{Var}_F(Y) &\approx \text{Var}_{\hat{F}}(Y) = \int (y-\mu)^2 d\hat{F}(y)
\end{aligned}$$

Approximation of Cumulative Distribution Function F

Expected value depends on F

$$
\begin{aligned}
\mathsf{Mean}_F(Y) &= E[Y] = \int y dF(y) \\
\mathsf{Var}_F(Y) &= E[(Y-\mu)^2] = \int (y-\mu)^2 dF(y) \\
\mathsf{h}_F(Y) &= E[h(Y)] = \int h(y) dF(y) \\
\mathsf{Mean}_F(Y) &\approx \mathsf{Mean}_{\hat{F}}(Y)) = \int y d\hat{F}(y) \\
\mathsf{Var}_F(Y) &\approx \mathsf{Var}_{\hat{F}}(Y) = \int (y-\mu)^2 d\hat{F}(y) \\
\mathsf{h}_F(Y) &\approx \mathsf{h}_{\hat{F}}(Y) = \int h(y) d\hat{F}(y)
\end{aligned}
$$

How to Take Expected Values Using $\hat{F}_N$?

How to Take Expected Values Using $\hat{F}_N$?

Simulation!

How to Take Expected Values Using $\hat{F}_N$?

Simulation!

Bootstrap is a double approximation:

How to Take Expected Values Using $\hat{F}_N$?

Simulation!
Bootstrap is a double approximation:

$$\mathsf{h}_F(Y)$$

How to Take Expected Values Using $\hat{F}_N$?

Simulation!
Bootstrap is a double approximation:

$$\mathsf{h}_F(Y) \approx \mathsf{h}_{\hat{F}}(Y)$$

How to Take Expected Values Using $\hat{F}_N$?

Simulation!

Bootstrap is a double approximation:

$$\mathsf{h}_F(Y) \approx \mathsf{h}_{\hat{F}}(Y) \approx \sum_{m=1}^{M} \frac{h(Y_m)}{M}$$

How to Take Expected Values Using $\hat{F}_N$?

Simulation!
Bootstrap is a double approximation:

$$\begin{aligned}
\mathsf{h}_F(Y) &\approx \mathsf{h}_{\hat{F}}(Y) \approx \sum_{m=1}^{M} \frac{h(Y_m)}{M} \\
h_F(\boldsymbol{\beta}^*) &\approx h_{\hat{F}}(\boldsymbol{\beta}^*) \approx \sum_{m=1}^{M} \frac{h(\boldsymbol{\beta}^*_m)}{M}
\end{aligned}$$

Percentile Bootstrap Confidence Intervals

- Suppose we have M Bootstrap iterations of our maximum likelihood estimator $\boldsymbol{\beta}^*$.
- Suppose $h(\boldsymbol{\beta}^*)$ is a scalar.
- Call $h(\boldsymbol{\beta}^*)_{\alpha}$ the value such that α of values less than or equal to it. (For example $h(\boldsymbol{\beta}^*)_{0.5}$ is the median (value such that 50% lower)
- A $1-\alpha$ confidence interval (under some additional assumptions) is

$$\mathsf{CI}_{1-\alpha} = \left(h(\boldsymbol{\beta}^*)_{\alpha/2}, h(\boldsymbol{\beta}^*)_{1-\alpha/2}\right)$$

Recall:

```
store_expected<- c()
for(z in 1:1000){
   subset<- sample(1:nrow(mips), nrow(mips), replace=T)
   use_mips<- mips[subset,]
   temp_reg<- glm(mip~treat + gop,
      family = binomial(link='probit'),
      data = as.data.frame(use_mips))
   store_expected[z]<- pnorm(temp_reg$coef%*%X_synth)
}
>conf_int_95<- quantile(store_expected, c(0.025, 0.975))
> conf_int_95
2.5% 97.5%
0.2953548 0.3546646
```

Bootstrapping is not a Panacea

Bootstrapping is not a Panacea

1) If you have dependence in your data, you need to sample to preserve it (time series, nested data) (but you can do a block bootstrap)

Bootstrapping is not a Panacea

1) If you have dependence in your data, you need to sample to preserve it (time series, nested data) (but you can do a block bootstrap)

Bootstrapping is not a Panacea

1) If you have dependence in your data, you need to sample to preserve it (time series, nested data) (but you can do a block bootstrap)
2) Bootstrap doesn't always work: (but you can fix it, usually)
   - Matching
   - LASSO
   - Maximum (e.g. MLE for uniform)

Inference Three Ways

1) Bootstrap $\rightsquigarrow$ (no asymptotics, simulation)
2) Delta Method $\rightsquigarrow$ (asymptotic normality, analytic)
3) Simulation from Multivariate Normal $\rightsquigarrow$ (asymptotic normality, simulation)

Delta Method (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*$.

(1)

Delta Method (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*$.
We are interested in $h(\boldsymbol{\beta}^*)$. (Assuming a large sample):

(1)

Delta Method (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*$.
We are interested in $h(\boldsymbol{\beta}^*)$. (Assuming a large sample):

$$h(\boldsymbol{\beta}^*) \quad \longrightarrow^{D} \quad \text{Multivariate Normal}\left(h(\boldsymbol{\beta}), \nabla h(\boldsymbol{\beta}^*)^{'} I_N(\boldsymbol{\beta}^*)^{-1} \nabla h(\boldsymbol{\beta}^*)\right) \tag{1}$$

Delta Method (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*$.
We are interested in $h(\boldsymbol{\beta}^*)$. (Assuming a large sample):

$$h(\boldsymbol{\beta}^*) \quad \longrightarrow^D \quad \text{Multivariate Normal}\left(h(\boldsymbol{\beta}), \nabla h(\boldsymbol{\beta}^*)' I_N(\boldsymbol{\beta}^*)^{-1} \nabla h(\boldsymbol{\beta}^*)\right) \tag{1}$$

Inference for $h(\boldsymbol{\beta}^*)$

Delta Method (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*$.
We are interested in $h(\boldsymbol{\beta}^*)$. (Assuming a large sample):

$$h(\boldsymbol{\beta}^*) \quad \longrightarrow^D \quad \text{Multivariate Normal}\left(h(\boldsymbol{\beta}), \nabla h(\boldsymbol{\beta}^*)' I_N(\boldsymbol{\beta}^*)^{-1} \nabla h(\boldsymbol{\beta}^*)\right)$$

Inference for $h(\boldsymbol{\beta}^*)$

1) Apply function h to maximum likelihood estimates:

(1)

Delta Method (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*$.
We are interested in $h(\boldsymbol{\beta}^*)$. (Assuming a large sample):

$$h(\boldsymbol{\beta}^*) \quad \longrightarrow^D \quad \text{Multivariate Normal} \left(h(\boldsymbol{\beta}), \nabla h(\boldsymbol{\beta}^*)^{'} I_N(\boldsymbol{\beta}^*)^{-1} \nabla h(\boldsymbol{\beta}^*) \right)$$

Inference for $h(\boldsymbol{\beta}^*)$

1) Apply function h to maximum likelihood estimates:

$$h(\boldsymbol{\beta}^*) \quad = \quad \Phi(\boldsymbol{X}_i^{'} \boldsymbol{\beta}^*) \tag{1}$$

Delta Method (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*$.
We are interested in $h(\boldsymbol{\beta}^*)$. (Assuming a large sample):

$$h(\boldsymbol{\beta}^*) \quad \longrightarrow^{D} \quad \text{Multivariate Normal}\left(h(\boldsymbol{\beta}), \nabla h(\boldsymbol{\beta}^*)^{'} I_N(\boldsymbol{\beta}^*)^{-1} \nabla h(\boldsymbol{\beta}^*)\right)$$

Inference for $h(\boldsymbol{\beta}^*)$

1) Apply function h to maximum likelihood estimates:

$$h(\boldsymbol{\beta}^*) \quad = \quad \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}^*) \tag{1}$$

2) To calculate variance-covariance matrix:

Delta Method (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*$.
We are interested in $h(\boldsymbol{\beta}^*)$. (Assuming a large sample):

$$h(\boldsymbol{\beta}^*) \quad \longrightarrow^D \quad \text{Multivariate Normal} \left(h(\boldsymbol{\beta}), \nabla h(\boldsymbol{\beta}^*)^{'} I_N(\boldsymbol{\beta}^*)^{-1} \nabla h(\boldsymbol{\beta}^*)\right)$$

Inference for $h(\boldsymbol{\beta}^*)$

1) Apply function h to maximum likelihood estimates:

$$h(\boldsymbol{\beta}^*) \quad = \quad \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}^*) \tag{1}$$

2) To calculate variance-covariance matrix:
   a) Calculate $\nabla h(\boldsymbol{\beta}^*)$ at maximum likelihood estimates:

Delta Method (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*$.
We are interested in $h(\boldsymbol{\beta}^*)$. (Assuming a large sample):

$$h(\boldsymbol{\beta}^*) \quad \longrightarrow^D \quad \text{Multivariate Normal}\left(h(\boldsymbol{\beta}), \nabla h(\boldsymbol{\beta}^*)^{'} I_N(\boldsymbol{\beta}^*)^{-1} \nabla h(\boldsymbol{\beta}^*)\right)$$

Inference for $h(\boldsymbol{\beta}^*)$

1) Apply function h to maximum likelihood estimates:

$$h(\boldsymbol{\beta}^*) \quad = \quad \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}^*)$$

2) To calculate variance-covariance matrix:
   a) Calculate $\nabla h(\boldsymbol{\beta}^*)$ at maximum likelihood estimates:

$$\nabla h(\boldsymbol{\beta}^*) \quad = \quad \left(\frac{\partial h(\boldsymbol{\beta})}{\partial \beta_1}|_{\boldsymbol{\beta}^*}, \frac{\partial h(\boldsymbol{\beta})}{\partial \beta_2}|_{\boldsymbol{\beta}^*}, \ldots, \frac{\partial h(\boldsymbol{\beta})}{\partial \beta_K}|_{\boldsymbol{\beta}^*}\right) \tag{1}$$

Delta Method (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*$.
We are interested in $h(\boldsymbol{\beta}^*)$. (Assuming a large sample):

$$h(\boldsymbol{\beta}^*) \quad \longrightarrow^D \quad \text{Multivariate Normal}\left(h(\boldsymbol{\beta}), \nabla h(\boldsymbol{\beta}^*)^{'} I_N(\boldsymbol{\beta}^*)^{-1} \nabla h(\boldsymbol{\beta}^*)\right)$$

Inference for $h(\boldsymbol{\beta}^*)$

1) Apply function h to maximum likelihood estimates:

$$h(\boldsymbol{\beta}^*) \quad = \quad \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}^*)$$

2) To calculate variance-covariance matrix:
   a) Calculate $\nabla h(\boldsymbol{\beta}^*)$ at maximum likelihood estimates:

$$\nabla h(\boldsymbol{\beta}^*) \quad = \quad \left(\frac{\partial h(\boldsymbol{\beta})}{\partial \beta_1}|_{\boldsymbol{\beta}^*}, \frac{\partial h(\boldsymbol{\beta})}{\partial \beta_2}|_{\boldsymbol{\beta}^*}, \ldots, \frac{\partial h(\boldsymbol{\beta})}{\partial \beta_K}|_{\boldsymbol{\beta}^*}\right)$$

   b) Calculate Variance covariance matrix:

(1)

Delta Method (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*$.
We are interested in $h(\boldsymbol{\beta}^*)$. (Assuming a large sample):

$$h(\boldsymbol{\beta}^*) \quad \longrightarrow^D \quad \text{Multivariate Normal}\left(h(\boldsymbol{\beta}), \nabla h(\boldsymbol{\beta}^*)^{'} I_N(\boldsymbol{\beta}^*)^{-1} \nabla h(\boldsymbol{\beta}^*)\right)$$

Inference for $h(\boldsymbol{\beta}^*)$

1) Apply function h to maximum likelihood estimates:

$$h(\boldsymbol{\beta}^*) \quad = \quad \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}^*)$$

2) To calculate variance-covariance matrix:
   a) Calculate $\nabla h(\boldsymbol{\beta}^*)$ at maximum likelihood estimates:

$$\nabla h(\boldsymbol{\beta}^*) \quad = \quad \left(\frac{\partial h(\boldsymbol{\beta})}{\partial \beta_1}|_{\boldsymbol{\beta}^*}, \frac{\partial h(\boldsymbol{\beta})}{\partial \beta_2}|_{\boldsymbol{\beta}^*}, \ldots, \frac{\partial h(\boldsymbol{\beta})}{\partial \beta_K}|_{\boldsymbol{\beta}^*}\right)$$

   b) Calculate Variance covariance matrix:

$$\nabla h(\boldsymbol{\beta}^*)^{'} I_N(\boldsymbol{\beta}^*)^{-1} \nabla h(\boldsymbol{\beta}^*)$$

Delta Method (Procedure, Then Justification)

Suppose we have maximum likelihood estimator $\boldsymbol{\beta}^*$.
We are interested in $h(\boldsymbol{\beta}^*)$. (Assuming a large sample):

$$h(\boldsymbol{\beta}^*) \quad \longrightarrow^D \quad \text{Multivariate Normal}\left(h(\boldsymbol{\beta}), \nabla h(\boldsymbol{\beta}^*)^{'} I_N(\boldsymbol{\beta}^*)^{-1} \nabla h(\boldsymbol{\beta}^*)\right)$$

Inference for $h(\boldsymbol{\beta}^*)$

1) Apply function h to maximum likelihood estimates:

$$h(\boldsymbol{\beta}^*) \quad = \quad \Phi(\boldsymbol{X}_i^{'}\boldsymbol{\beta}^*)$$

2) To calculate variance-covariance matrix:
   a) Calculate $\nabla h(\boldsymbol{\beta}^*)$ at maximum likelihood estimates:

$$\nabla h(\boldsymbol{\beta}^*) \quad = \quad \left(\frac{\partial h(\boldsymbol{\beta})}{\partial \beta_1}|_{\boldsymbol{\beta}^*}, \frac{\partial h(\boldsymbol{\beta})}{\partial \beta_2}|_{\boldsymbol{\beta}^*}, \ldots, \frac{\partial h(\boldsymbol{\beta})}{\partial \beta_K}|_{\boldsymbol{\beta}^*}\right)$$

   b) Calculate Variance covariance matrix:

$$\nabla h(\boldsymbol{\beta}^*)^{'}(-E[\text{Hessian}]^{-1})\nabla h(\boldsymbol{\beta}^*)$$

Example:

$$\frac{\partial E[Y|\boldsymbol{X}_i, \boldsymbol{\beta}]}{\partial \boldsymbol{\beta}_k} = \phi(\boldsymbol{X}_i \boldsymbol{\beta}) X_k$$

```
prob_grad<- function(X, coef){
     base<- dnorm(coef%*%X)
     out<- base*X
     return(out)
   }
> mle_grad<- prob_grad(X_synth, reg1$coef)

> mle_grad
[1] 0.359335 0.000000 0.359335

> ses<- sqrt(t(mle_grad)%*%vcov(reg1)%*%mle_grad)

> ses
[1,] 0.01618305
```

Comparison

	Bootstrap	Delta Method
Standard Error	0.016	0.016
Confidence Interval	[0.295, 0.355]	

Delta Method (Justification)

Recall that

Delta Method (Justification)

Recall that

$$\boldsymbol{\beta}^* \quad \longrightarrow^D \quad \text{Multivariate Normal}\left(\boldsymbol{\beta}, I_N(\boldsymbol{\beta}^*)^{-1}\right)$$

Delta Method (Justification)

Recall that

$$\boldsymbol{\beta}^* \quad \longrightarrow^D \quad \text{Multivariate Normal}\left(\boldsymbol{\beta}, I_N(\boldsymbol{\beta}^*)^{-1}\right)$$

Multivariate Taylor Series:

Delta Method (Justification)

Recall that

$$\boldsymbol{\beta}^* \quad \longrightarrow^D \quad \text{Multivariate Normal}\left(\boldsymbol{\beta}, I_N(\boldsymbol{\beta}^*)^{-1}\right)$$

Multivariate Taylor Series:

$$h(\boldsymbol{\beta}^*) \quad = \quad h(\boldsymbol{\beta}) + \nabla h(\boldsymbol{\beta})^{'}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) + \text{ Remainder}$$

Delta Method (Justification)

Recall that

$$\boldsymbol{\beta}^* \quad \longrightarrow^{D} \quad \text{Multivariate Normal}\left(\boldsymbol{\beta}, I_N(\boldsymbol{\beta}^*)^{-1}\right)$$

Multivariate Taylor Series:

$$h(\boldsymbol{\beta}^*) \quad = \quad h(\boldsymbol{\beta}) + \nabla h(\boldsymbol{\beta})^{'}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) + \text{ Remainder}$$

Delta Method (Justification)

Take the limit:

$$
\begin{aligned}
\lim_{n\rightarrow\infty} h(\boldsymbol{\beta}^{*}) \quad &= \quad \lim_{n\rightarrow\infty}\left(h(\boldsymbol{\beta}) + \nabla h(\boldsymbol{\beta})^{'}(\boldsymbol{\beta}^{*}-\boldsymbol{\beta}) + \text{ Remainder }\right)\\
\lim_{n\rightarrow\infty} h(\boldsymbol{\beta}^{*}) \quad &\longrightarrow^{P} \quad h(\boldsymbol{\beta})\\
\lim_{n\rightarrow\infty} h(\boldsymbol{\beta}^{*}) \quad &\longrightarrow^{D} \quad h(\boldsymbol{\beta}) + \lim_{n\rightarrow\infty}\left(\nabla h(\boldsymbol{\beta})^{'} \quad \underbrace{(\boldsymbol{\beta}^{*}-\boldsymbol{\beta})}_{\text{MVN}(0, I_N(\boldsymbol{\beta})^{-1}}\right)
\end{aligned}
$$

Delta Method (Justification)

Take the limit:

$$\begin{aligned}
\lim_{n\rightarrow\infty} h(\boldsymbol{\beta}^{*}) &= \lim_{n\rightarrow\infty}\left(h(\boldsymbol{\beta}) + \nabla h(\boldsymbol{\beta})^{'}(\boldsymbol{\beta}^{*} - \boldsymbol{\beta}) + \text{ Remainder }\right)\\
\lim_{n\rightarrow\infty} h(\boldsymbol{\beta}^{*}) &\longrightarrow^{P} h(\boldsymbol{\beta})\\
\lim_{n\rightarrow\infty} h(\boldsymbol{\beta}^{*}) &\longrightarrow^{D} h(\boldsymbol{\beta}) + \lim_{n\rightarrow\infty}\left(\nabla h(\boldsymbol{\beta})^{'} \underbrace{(\boldsymbol{\beta}^{*} - \boldsymbol{\beta})}_{\text{MVN}(0, I_N(\boldsymbol{\beta})^{-1}}\right)
\end{aligned}$$

Then by Slutsky's Theorem:

$$h(\boldsymbol{\beta}^{*}) \longrightarrow^{D} \text{MVN}(h(\boldsymbol{\beta}), \nabla h(\boldsymbol{\beta}^{*})^{'} I_N(\boldsymbol{\beta}^{*})^{-1} \nabla h(\boldsymbol{\beta}^{*}))$$

Delta Method Confidence Intervals

- Suppose $h(\boldsymbol{\beta}^*)$ is a scalar
- $\mathsf{SE}(h(\boldsymbol{\beta}^*)) = \sqrt{(\nabla h(\boldsymbol{\beta}^*)' I_N(\boldsymbol{\beta}^*)^{-1} \nabla h(\boldsymbol{\beta}^*))}$
- Then a $1-\alpha$ confidence interval is:

$$\mathsf{CI}_{1-\alpha} = [h(\boldsymbol{\beta}^*) - 1.96 \times \mathsf{SE}(h(\boldsymbol{\beta}^*)), h(\boldsymbol{\beta}^*) + 1.96 \times \mathsf{SE}(h(\boldsymbol{\beta}^*))]$$

Delta Method: Example

```
> y.prob<- pnorm(y.tilde)
> X_synth<- c(1, 0, 1)
> y.tilde<- reg1$coef%*%X_synth
> y.prob<- pnorm(y.tilde)
> y.prob
[1,] 0.3237277

> ses<- sqrt(t(mle_grad)%*%vcov(reg1)%*%mle_grad)
> ses
[1,] 0.01618305
> delta_95<- c(y.prob - 1.96*ses, y.prob + 1.96*ses)
> delta_95
[1] 0.2920089 0.3554464
```

Comparison

	Bootstrap	Delta Method
Standard Error	0.016	0.016
Confidence Interval	[0.295, 0.355]	[0.292, 0.355]

Inference Three Ways

1) Bootstrap $\rightsquigarrow$ (no asymptotics, simulation)
2) Delta Method $\rightsquigarrow$ (asymptotic normality, analytic)
3) Simulation from Multivariate Normal $\rightsquigarrow$ (asymptotic normality, simulation)

Simulation from Multivariate Normal (Procedure)

$\boldsymbol{\beta}^* \longrightarrow^D \mathsf{MVN}(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1})$

Simulation from Multivariate Normal (Procedure)

$\boldsymbol{\beta}^* \longrightarrow^D$ MVN$(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1})$
Inference about $h(\boldsymbol{\beta}^*)$:

Simulation from Multivariate Normal (Procedure)

$\boldsymbol{\beta}^* \longrightarrow^D \text{MVN}(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1})$
Inference about $h(\boldsymbol{\beta}^*)$:
For each simulation m of M simulations:

Simulation from Multivariate Normal (Procedure)

$\boldsymbol{\beta}^* \longrightarrow^D \text{MVN}(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1})$
Inference about $h(\boldsymbol{\beta}^*)$:
For each simulation m of M simulations:

$$\boldsymbol{\beta}^m \quad \sim \quad \text{MVN}\left(\boldsymbol{\beta}^*, I_N(\boldsymbol{\beta})^{-1}\right)$$

Simulation from Multivariate Normal (Procedure)

$\boldsymbol{\beta}^* \longrightarrow^D \textsf{MVN}(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1})$
Inference about $h(\boldsymbol{\beta}^*)$:
For each simulation m of M simulations:

$$\begin{aligned} \boldsymbol{\beta}^m &\sim \textsf{MVN}\left(\boldsymbol{\beta}^*, I_N(\boldsymbol{\beta})^{-1}\right) \\ h(\boldsymbol{\beta})^m &= h(\boldsymbol{\beta}^m) \end{aligned}$$

Simulation from Multivariate Normal (Procedure)

$\boldsymbol{\beta}^* \longrightarrow^D$ MVN$(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1})$
Inference about $h(\boldsymbol{\beta}^*)$:
For each simulation m of M simulations:

$$\begin{aligned} \boldsymbol{\beta}^m & \sim \text{MVN}\left(\boldsymbol{\beta}^*, I_N(\boldsymbol{\beta})^{-1}\right) \\ h(\boldsymbol{\beta})^m & = h(\boldsymbol{\beta}^m) \end{aligned}$$

Use collection of simulated values $\left(h(\boldsymbol{\beta})^1, h(\boldsymbol{\beta})^2, \ldots, h(\boldsymbol{\beta})^M\right)$ to approximate distribution of $h(\boldsymbol{\beta}^*)$.

Simulation from Multivariate Normal (Procedure)

$\boldsymbol{\beta}^* \longrightarrow^D \text{MVN}(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1})$
Inference about $h(\boldsymbol{\beta}^*)$:
For each simulation m of M simulations:

$$\begin{aligned}
\boldsymbol{\beta}^m &\sim \text{MVN}\left(\boldsymbol{\beta}^*, I_N(\boldsymbol{\beta})^{-1}\right) \\
h(\boldsymbol{\beta})^m &= h(\boldsymbol{\beta}^m)
\end{aligned}$$

Use collection of simulated values $\left(h(\boldsymbol{\beta})^1, h(\boldsymbol{\beta})^2, \ldots, h(\boldsymbol{\beta})^M\right)$ to approximate distribution of $h(\boldsymbol{\beta}^*)$.

$$E[h(\boldsymbol{\beta}^*)] = \sum_{m=1}^{M} \frac{h(\boldsymbol{\beta})^m}{M}$$

Simultation from Multivariate Normal (Procedure)

$\boldsymbol{\beta}^* \longrightarrow^D \text{MVN}(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1})$
Inference about $h(\boldsymbol{\beta}^*)$:
For each simulation m of M simulations:

$$\begin{aligned} \boldsymbol{\beta}^m &\sim \text{MVN}\left(\boldsymbol{\beta}^*, I_N(\boldsymbol{\beta})^{-1}\right) \\ h(\boldsymbol{\beta})^m &= h(\boldsymbol{\beta}^m) \end{aligned}$$

Use collection of simulated values $\left(h(\boldsymbol{\beta})^1, h(\boldsymbol{\beta})^2, \ldots, h(\boldsymbol{\beta})^M\right)$ to approximate distribution of $h(\boldsymbol{\beta}^*)$.

$$\text{var}[h(\boldsymbol{\beta}^*)] = \sum_{m=1}^{M} \frac{\left(h(\boldsymbol{\beta})^m - \sum_{m=1}^{M} \frac{h(\boldsymbol{\beta})^m}{M}\right)^2}{M}$$

Simulation from Multivariate Normal (Procedure)

$\boldsymbol{\beta}^* \longrightarrow^D \text{MVN}(\boldsymbol{\beta}, I(\boldsymbol{\beta})^{-1})$
Inference about $h(\boldsymbol{\beta}^*)$:
For each simulation m of M simulations:

$$\begin{aligned} \boldsymbol{\beta}^m &\sim \text{MVN}\left(\boldsymbol{\beta}^*, I_N(\boldsymbol{\beta})^{-1}\right) \\ h(\boldsymbol{\beta})^m &= h(\boldsymbol{\beta}^m) \end{aligned}$$

Use collection of simulated values $\left(h(\boldsymbol{\beta})^1, h(\boldsymbol{\beta})^2, \ldots, h(\boldsymbol{\beta})^M\right)$ to approximate distribution of $h(\boldsymbol{\beta}^*)$.

$$f[h(\boldsymbol{\beta}^*)] = \sum_{m=1}^{M} \frac{f(h(\boldsymbol{\beta})^m)}{M}$$

Simulation: Example

```
libary(MASS)
draw_coef<- mvrnorm(1000, mu = reg1$coef, Sigma = vcov(reg1))

> dim(draw_coef)
[1] 1000 3

dist_exp<- pnorm(draw_coef%*%X_synth)
> length(dist_exp)
[1] 1000

sd(dist_exp)
[1] 0.01572543
```

Comparison

	Bootstrap	Delta Method	MVN Simulation
Standard Error	0.016	0.016	0.016
Confidence Interval	[0.295, 0.355]	[0.292, 0.355]	

Simultation from Multivariate Normal (Justification)

Simulation is a (stronger assumption) double approximation

$$\begin{aligned} E_f[h(\boldsymbol{\beta})^*] &= \int h(\boldsymbol{\beta})^* f(\boldsymbol{\beta}^*) d\boldsymbol{\beta}^* \\ &\approx \int h(\boldsymbol{\beta}^*) \underbrace{\tilde{f}(\boldsymbol{\beta}^*)}_{\text{MVN}} d\boldsymbol{\beta}^* \approx \sum_{m=1}^{M} \frac{h(\boldsymbol{\beta})^m}{M} \end{aligned}$$

Percentile Simulation Confidence Intervals

- Suppose we have M draws
- Suppose $h(\boldsymbol{\beta}^*)$ is a scalar
- Then a $1 - \alpha$ confidence interval is:

$$\mathsf{CI}_{1-\alpha} = \left[h(\boldsymbol{\beta}^*)_{\alpha/2}, h(\boldsymbol{\beta}^*)_{1-\alpha/2}\right]$$

- In practice: order the values, select quantile values at $\alpha/2$ and $1 - \alpha/2$.

Simulation Confidence Intervals

```
sim_95<- quantile(dist_exp, c(0.025, 0.975))
> sim_95
2.5% 97.5%
0.2953568 0.3559635
```

Comparison

	Bootstrap	Delta Method	MVN Simulation
Standard Error	0.016	0.016	0.016
Confidence Interval	[0.295, 0.355]	[0.292, 0.355]	[0.295, 0.356]

So Which Should I Use?

Lots of data

- Then all three methods will be close
- Delta Method + MVN Simulation $\rightsquigarrow$ faster (but what else are you doing?)

Not lots of data

- Then only bootstrap will perform well
- And will be comparable in speed to Delta + MVN

Thursday

Model fit for Binary DVs + Ordered Probit