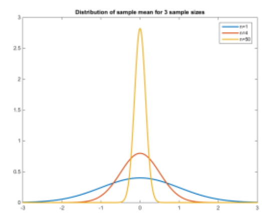
Convergence

 1

Distribution of sample mean

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$



 2

Question for class

Draw a sample of size $n = 2$ of uniform(0,1) independent random variables and compute the sample mean. Do this a lot of times and compute the sample mean. Plot the empirical distribution. Now do it again with $n = 1{,}000$. Does either distribution look familiar?

 3

Sequence of random variables

$$x_1, x_2 \ldots$$

Might be a statistic from samples of size 1,2, …—or maybe not.

x_i , x_j might be independent—or maybe not.

x_i , x_j might be distributed identically—or maybe not.

 6

No limit

Would be nice if

$$\lim_{n\to\infty} \bar{x} = \mu$$

Unfortunately, no limiting argument exists.

Convergence in probability

Definition 5.5.1 A sequence of random variables $x_1, x_2, \ldots$, *converges in probability* to a random variable x if $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(|x_n - x| \geq \varepsilon) = 0$$

or equivalently

$$\lim_{n\to\infty} P(|x_n - x| < \varepsilon) = 1$$

Consistency and probability limit

If x_n is a sequence of estimators that converges to a constant x, we say that x_n is a *consistent* estimator of x.

When x is a constant, we abbreviate the convergence argument by saying x is the *probability limit* of x_n or

$$\operatorname{plim} x_n = x$$

$$x_n \xrightarrow{p} x$$

Weak Law of Large Numbers

Theorem 5.5.2 (Weak Law of Large Numbers)

Let $x_1, x_2, \ldots$ be iid random variables with $\mathrm{E}(x_i) = \mu$ and $\operatorname{var}(x_i) = \sigma^2 < \infty$. Then $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(|\bar{x}_n - \mu| < \varepsilon) = 1$$

$$\operatorname{plim}(\bar{x}_n) = \mu$$

$\bar{x}_n$ converges in probability to μ.

Proof of WLLN

Reminder of Chebychev's Inequality

$$P(g(x) \geq r) \leq \frac{\mathrm{E}\big(g(x)\big)}{r}$$

$$P(|\bar{x}_n - \mu| \geq \varepsilon) = P\big((\bar{x}_n - \mu)^2 \geq \varepsilon^2\big)$$
$$\leq \frac{\mathrm{E}[(\bar{x}_n - \mu)^2]}{\varepsilon^2} = \frac{\mathrm{var}(\bar{x}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

Which goes to zero as $n \to \infty$.

WLLN for s^2

$$P(|s_n^2 - \sigma^2| \geq \varepsilon) \leq \frac{\mathrm{E}[(s_n^2 - \sigma^2)^2]}{\varepsilon^2} = \frac{\mathrm{var}(s_n^2)}{\varepsilon^2}$$

We know that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$. So

$$\mathrm{var}(s_n^2) = 2(n-1)\left[\frac{\sigma^2}{n-1}\right]^2 = \frac{2\sigma^4}{n-1}$$

WLLN for moments

$$\bar{m}'_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

$$\mathrm{E}(\bar{m}'_k) = \mu'_k$$

And

$$\mathrm{plim}(\bar{m}'_k) = \mu'_k$$

plims for continuous functions

Theorem 5.5.4 Suppose that $x_1, x_2, \ldots$ converges in probability to x and that h is a continuous function. Then $h(x_1), h(x_2), \ldots$ converges in probability to $h(x)$.

Or

If h is continuous

$$\mathrm{plim}(x_n) = x \Rightarrow \mathrm{plim}\big(h(x_n)\big) = h(x)$$

Example: $s = \sqrt{s^2}$

$$\mathrm{E}(s) = ? \neq \sigma$$
$$\mathrm{plim}(s) = \sigma$$

Unbiased vs Consistent

s is a consistent but biased estimator of σ.

Let $u \sim N(0, v)$,
$\bar{x} + u$ is an unbiased but inconsistent estimator of μ.

Asymptotically unbiased

If

$$\lim_{n\to\infty} \mathrm{E}(x_n) - x = 0$$

then x_n is asymptotically unbiased.

Note that consistency is usually sufficient to imply asymptotically unbiased.

Silly counter-example

$$x_n = \begin{cases} \mu, p = \dfrac{n-1}{n} \\ \mu + n, p = \dfrac{1}{n} \end{cases}$$

$$\mathrm{plim}(x_n) = \mu$$

$$\mathrm{E}(x_n) = \mu + 1$$

Almost sure convergence (or convergence with probability 1)

A sequence of random variables $x_1, x_2, \ldots$ *converges almost surely* to a random variable x if $\forall \varepsilon > 0$,

$$P\left(\lim_{n\to\infty} |x - x_n| < \varepsilon\right) = 1$$

$$x_n \xrightarrow{a.s.} x$$

Almost sure convergence example

Example: Suppose $s \sim U[0,1]$ and consider the odd sequence

$$X_n(s) = s + s^n$$

On the open interval $[0,1)$ $s^n \to 0$ so $X_n(s)$ converges to the variable $s \sim U[0,1)$. But there is an infinitesimal chance of $s = 1$, which gives $X_n(1) = 2$. So there is almost sure convergence

Almost surely⇒plim

Convergence almost surely implies convergence in probability.

Strong Law of Large Numbers

Theorem 5.5.9 (Strong Law of Large Numbers)
Let $x_1, x_2, \ldots$ be iid random variables with $\mathrm{E}(x_i) = \mu$ and $\mathrm{var}(x_i) = \sigma^2 < \infty$. Then $\forall \varepsilon > 0$,

$$P\left(\lim_{n\to\infty} |\bar{x}_n - \mu| < \varepsilon\right) = 1$$

$\bar{x}_n$ converges almost surely to μ.

plim for mean of correlated sample

$$x_t = \mu + \varepsilon_t$$
$$\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t$$
$$|\rho| < 1, \eta_t \sim iid(0, \sigma^2 < \infty), \text{plim}\, \overline{x_t} = \text{plim}\, \overline{x_{t-1}} = \alpha$$

$$\hat{\beta} = \frac{1}{n}\sum(x_t - \rho x_{t-1})$$

$$x_t - \rho x_{t-1} = (\mu + \varepsilon_t) - \rho(\mu + \varepsilon_{t-1}) = (1-\rho)\mu + \eta_t$$

The η_t are iid, so by *lln* $\text{plim}\, \hat{\beta} = (1-\rho)\mu$

$$\hat{\beta} = \overline{x_t} - \rho\overline{x_{t-1}}$$
$$\text{plim}\, \hat{\beta} = \text{plim}\, \overline{x_t} - \rho\, \text{plim}\, \overline{x_{t-1}}$$
$$(1-\rho)\mu = \alpha - \rho\alpha$$
$$\text{plim}\, \overline{x_t} = \mu$$

Finite sample vs consistent

Consider the estimator

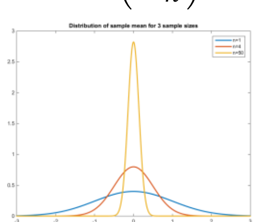
$$\hat{\mu} = \mu \pm 0.1, with \frac{50}{50} odds$$

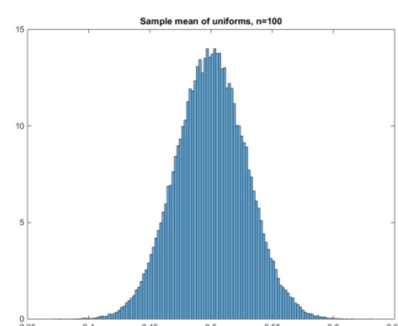
Mean square error is 0.01.

Compare to sample mean of iid normals with mean square error $\frac{\sigma^2}{n}$. So is n 100 times larger than σ^2?

Distribution of sample mean

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$





Question for class

The mean, $\bar{x}$, of n Bernoulli trials with $p = 1/2$ is distributed approximately $\bar{x} \overset{A}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$. That suggests that 10 percent of the time the mean should turn out to be greater than $p + 1.2816 \times \sqrt{p(1-p)/n}$. Run a series of Monte Carlo simulations to find out how good an approximation this is for a variety of values of n.

Convergence in distribution

A sequence of random variables $x_1, x_2, \ldots$ *converges in distribution* to a random variable X if

$$\lim_{n \to \infty} F_{x_n}(x) = F_X(x)$$

at all points where $F_X(x)$ is continuous.

Or

$$x_n \overset{d}{\rightarrow} X$$
$$x_n \overset{A}{\sim} F_X(x)$$

Convergence in probability implies convergence in distribution.

$$x_n \overset{p}{\rightarrow} X \Rightarrow x_n \overset{d}{\rightarrow} X$$

Maximum of n iid $U(0,1)$

Suppose $x_1, x_2, \ldots$ are iid uniform(0,1) and we're interested in

$$x_n = \max_{1 \le i \le n} x_i$$

$$\varepsilon > 0,$$
$$F_{X_n}(x_n) = P(x_n \le 1 - \varepsilon) = (1 - \varepsilon)^n$$

$$\text{plim}(x_n) = 1$$

Maximum of n iid $U(0,1)$

$$F_{X_n}(x_n) = P(x_n \le 1 - \varepsilon) = (1-\varepsilon)^n$$
$$\varepsilon = t/n$$

$$P(x_n \le 1 - t/n) = (1 - t/n)^n$$
$$\lim_{n\to\infty} (1 - t/n)^n = e^{-t}$$

for large n we have

$$P(x_n \le 1 - t/n) = e^{-t}$$

Maximum of n iid $U(0,1)$

$$P\left(x_n \le 1 - \frac{t}{n}\right) = P(n(x_n - 1) \le -t)$$
$$= P(n(1 - x_n) \le t) \to 1 - e^{-t}$$

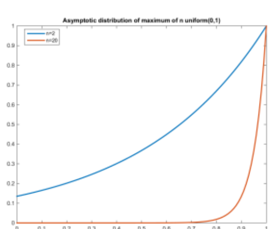
$$P(n(x_n - 1) \le -t) = P(n(1-x_n) \ge t) = e^{-t}$$
$$= P(n(1 - x_n) \le t) \to 1 - e^{-t}$$
$$n(1 - x_n) \overset{A}{\sim} \exp(1)$$

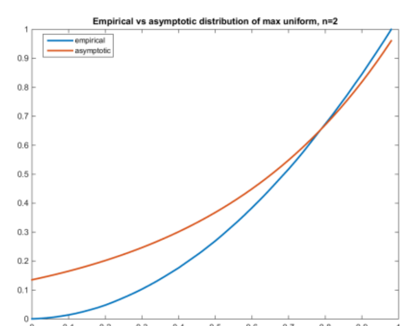
If we use the change of variable formula we get

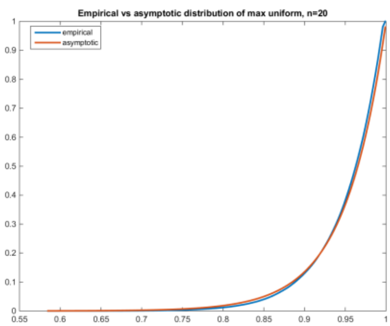
$$x_n \overset{A}{\to} e^{-n(1-x)}$$

Asymptotic distribution

$$x_n \overset{A}{\to} e^{-n(1-x)}$$







 37

Central Limit Theorem

Central Limit Theorem (CLT) (Lindeberg-Lévy): Let $x_1, x_2, \ldots$ be a sequence of iid random variables whose moment generating functions exist in a neighborhood of 0 (finite moments). Let $\mathrm{E}(x_i) = \mu, \mathrm{var}(x_i) = \sigma^2 > 0$. Then

$$\frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma} \xrightarrow{d} \Phi(x_n)$$

In other words, the standardized sample mean is asymptotically standard normal.

 38

Or

$$\frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma} \xrightarrow{d} \Phi(x_n)$$

$$\sqrt{n}(\bar{x}_n - \mu) \overset{A}{\sim} N(0, \sigma^2)$$

$$\bar{x}_n \overset{A}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

 39

CLT for noncentral moments

Let $x_1, x_2, \ldots$ be a sequence of iid random variables whose moment generating functions exist in a neighborhood of 0 (finite moments). Let $\mathrm{E}(x_i) = \mu, \mathrm{var}(x_i) = \sigma^2 > 0$. Then

$$\sqrt{n}(\overline{m}'_k - \mu'_k) \xrightarrow{d} N\big(0, \mu'_{2k} - \mu'^2_k\big)$$

 40

Central limit theorem for stationary stochastic processes

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$

where ε_t is i.i.d. and $\sigma_\varepsilon^2 < \infty$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$

This is called an infinite order *moving average* process.

We also define the covariances

$$\gamma_j = \mathrm{E}\big[(y_t - \mu)\big(y_{t-j} - \mu\big)\big]$$

CLT for stationary stochastic processes

$$\sqrt{T}(\bar{y} - \mu) \overset{A}{\sim} N\left(0, \sum_{j=-\infty}^{\infty} \gamma_j\right)$$

First-order autoregressive

$$y_t = (1-\rho)\mu + \rho y_{t-1} + \varepsilon_t, |\rho| < 1$$
$$y_t = \varepsilon_t + (1-\rho)\mu + \rho y_{t-1}$$
$$= \varepsilon_t + (1-\rho)\mu + \rho[\varepsilon_{t-1} + (1-\rho)\mu + \rho y_{t-2}]$$
$$= \varepsilon + \rho_{\varepsilon_{t-1}} + (1-\rho)\mu[1+\rho]$$
$$+\rho^2[\varepsilon_{t-2} + (1-\rho)\mu + \rho y_{t-3}]$$

Or, collecting terms

$$y_t = \mu + \ \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \cdots$$

$$(y_t - \mu)^2 = \varepsilon_t^2 + \rho^2\varepsilon_{t-1}^2 + \rho^{2^2}\varepsilon_{t-2}^2 + \cdots$$
$$\gamma_0 = \frac{1}{1-\rho^2}\sigma_\varepsilon^2$$

$$(y_t - \mu)(y_{t-1} - \mu) =$$
$$\varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \cdots$$
$$\times \qquad \varepsilon_{t-1} + \rho\varepsilon_{t-2} + \rho^2\varepsilon_{t-3} + \cdots$$
$$= \rho\varepsilon_{t-1}^2 + \rho^3\varepsilon_{t-2}^2 + \cdots$$

$$\gamma_1 = \rho\gamma_0$$
$$\gamma_j = \rho^j\gamma_0$$

$$\gamma_j = \rho^j \gamma_0$$
$$\sum_{j=1}^{\infty} \gamma_j = \gamma_0[\rho + \rho^2 + \rho^3 + \cdots] = \gamma_0 \frac{\rho}{1-\rho}$$
$$\sum_{j=-\infty}^{\infty} \gamma_j = \gamma_0 \frac{\rho}{1-\rho} + \gamma_0 + \gamma_0 \frac{\rho}{1-\rho}$$
$$= \gamma_0 \left[\frac{\rho + (1-\rho) + \rho}{1-\rho}\right]$$
$$= \frac{1}{1-\rho^2} \sigma_\varepsilon^2 \times \frac{1+\rho}{1-\rho} = \frac{\sigma_\varepsilon^2}{(1-\rho)^2}$$
$$\sqrt{T}(\bar{y} - \mu) \overset{A}{\sim} N\left(0, \frac{\sigma_\varepsilon^2}{(1-\rho)^2}\right)$$

Slutsky's Theorem

Slutsky's Theorem: If $X_n \overset{d}{\to} X$ and $Y_n \overset{p}{\to} a$, a constant, then

a. $Y_n X_n \overset{d}{\to} aX$

b. $Y_n + X_n \overset{d}{\to} a + X$

Slutsky example

$$z = \frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma} \overset{d}{\to} N(0,1)$$

$$t = \frac{\sqrt{n}(\bar{x}_n - \mu)}{s_n} = \frac{\sigma}{s_n} \times \frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma} \overset{d}{\to} N(0,1)$$

Taylor series

If a function $g(x)$ has derivatives of order r, $g^{(r)}(x)$, then for any constant a the r^{th} order *Taylor series approximation* around a is

$$T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(x-a)^i$$

Example

$$T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(x-a)^i$$

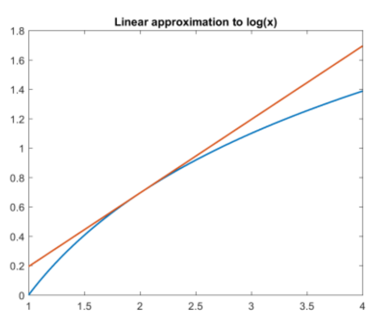
For example, if

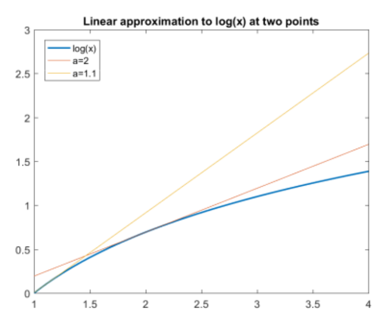
$$g(x) = 3 - x^2$$
$$a = 0$$
$$T_0(x) = \frac{3}{1}(x-0)^0 = 3$$
$$T_1(x) = 3 + \frac{-2 \times 0}{1}(x-0)^1 = 3$$
$$T_2(x) = 3 + \frac{-2}{1 \times 2}(x-0)^2 = 3 - x^2$$
$$T_{r>2}(x) = 3 - x^2$$

Linear approximation

$$y = \log(x)$$

$$y \approx \frac{\log(a)}{1} + \frac{1}{a}(x-a)$$





Variance of a nonlinear function

Suppose we have a vector of random variables $T = (T_1, \dots, T_k)$ with means $\theta = (\theta_1, \dots, \theta_k)$.

$$g(t) \approx g(\theta) + \sum_{i=1}^{k} \left.\frac{\partial g}{\partial t_i}\right|_{\theta} (t_i - \theta_i)$$

$$\mathrm{E}\big(g(t)\big) \approx g(\theta) + \sum_{i=1}^{k} \left.\frac{\partial g}{\partial t_i}\right|_{\theta} \mathrm{E}(t_i - \theta_i) = g(\theta)$$

Variance of a nonlinear function

$$g(t) \approx g(\theta) + \sum_{i=1}^{k} \left.\frac{\partial g}{\partial t_i}\right|_{\theta} (t_i - \theta_i)$$

$$\mathrm{var}\big(g(T)\big) \approx \mathrm{E}([g(T) - g(\theta)]^2)$$

$$\approx \mathrm{E}\left(\left[\sum_{i=1}^{k} \left.\frac{\partial g}{\partial t_i}\right|_{\theta} (t_i - \theta_i)\right]^2\right)$$

$$= \sum_{i=1}^{k} \left[\left.\frac{\partial g}{\partial t_i}\right|_{\theta}\right]^2 \mathrm{var}(t_i) + 2\sum_{i>j} \left.\frac{\partial g}{\partial t_i}\right|_{\theta} \left.\frac{\partial g}{\partial t_j}\right|_{\theta} \mathrm{cov}(t_i, t_j)$$

Matrix version

$$g_{i,j}^{(1)} = \partial g_i / \partial T_j$$

$$\mathrm{var}\big(g(t)\big) \approx g_{i,j}^{(1)}\, \mathrm{var}(T) g_{i,j}^{(1)\prime}$$

where $\mathrm{var}(T) = \mathrm{E}[(T - \theta)(T - \theta)']$

Scalar example

$$\theta \sim (\bar{\theta}, \sigma^2)$$

$$\mathrm{var}(log(\theta)) = ?$$

$$\frac{d \log(\theta)}{d\theta} = \frac{1}{\theta}$$

$$\mathrm{var}(\log(\theta)) \approx \left[\left.\frac{d \log(\theta)}{d\theta}\right|_{\bar{\theta}}\right]^2 \mathrm{var}(\theta) = \frac{\sigma^2}{\bar{\theta}^2}$$

Question for class

Suppose $x \sim U(0, u)$. Consider the function $g(\theta) = \log(\theta)$. It turns out that the first-order Taylor series approximation to the variance of a function of a random variable is

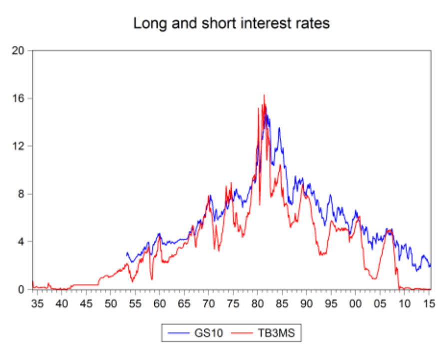
$$\left[\frac{dg(\theta)}{d\theta}_{|\mathrm{E}(\theta))}\right]^2 \mathrm{var}(\theta)$$

Use this expression to calculate and approximate variance of $\log(\theta)$ for $u = 1$ and for $u = 0.01$. Now redo the calculation by generating simulated $\theta's$ and finding the variance of $\log(\theta)$.

Delta method

Let Y_n be a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$. Then

$$\sqrt{n}\big(g(Y_n) - g(\theta)\big) \xrightarrow{d} N(0, \sigma^2[g'(\theta)]^2)$$



Interest rate example

$$i_t^{3m} = \beta_1 + \beta_2 i_t^{10y} + \beta_3 i_{t-1}^{3m} + \varepsilon_t$$

Effect of a maintained unit change in i_t^{10y}.

$$\beta_2$$
$$\beta_2 + \beta_3\beta_2$$
$$\vdots$$
$$\frac{\beta_2}{1 - \beta_3}$$

$$g(\beta_2,\beta_3)=\frac{\beta_2}{1-\beta_3}$$
$$g'=\left[\frac{1}{1-\beta_3}\quad -\frac{\beta_2}{(1-\beta_3)^2}\right]=\frac{1}{1-\beta_3}[1\quad -g]$$

$$\operatorname{var}\left(\frac{\beta_2}{1-\beta_3}\right)$$
$$\approx\left(\frac{1}{1-\beta_3}\right)^2[1\quad -g]\begin{bmatrix}\operatorname{var}(\beta_2) & \operatorname{cov}(\beta_2,\beta_3)\\ \operatorname{cov}(\beta_2,\beta_3) & \operatorname{var}(\beta_3)\end{bmatrix}\begin{bmatrix}1\\ -g\end{bmatrix}$$

Dependent Variable: TB3MS
Method: Least Squares
Date: 07/04/15 Time: 10:36
Sample (adjusted): 1953M04 2015M05
Included observations: 746 after adjustments

Variable	Coefficien...	Std. Error	t-Statistic	Prob.
C	-0.113023	0.041044	-2.753670	0.0060
GS10	0.067707	0.013948	4.854201	0.0000
TB3MS(-1)	0.934482	0.012752	73.28215	0.0000

R-squared	0.981629	Mean dependent var	4.534357
Adjusted R-squared	0.981579	S.D. dependent var	3.064075
S.E. of regression	0.415866	Akaike info criterion	1.087105
Sum squared resid	128.4977	Schwarz criterion	1.105663
Log likelihood	-402.4903	Hannan-Quinn criter.	1.094258
F-statistic	19850.25	Durbin-Watson stat	1.249119
Prob(F-statistic)	0.000000		

$$\hat{g}=\frac{\hat{\beta}_2}{1-\hat{\beta}_3}=1.034$$

Estimated variance-covariance matrix

	C	GS10	TB3MS(-1)
C	0.001685	-0.000428	0.000248
GS10	-0.000428	0.000195	-0.000164
TB3MS(-1)	0.000248	-0.000164	0.000163

Asymptotic estimate of asymptotic effect

$$\operatorname{var}\left(\frac{\beta_2}{1-\beta_3}\right)$$
$$\approx\left(\frac{1}{1-.934}\right)^2[1\quad -1.034]\begin{bmatrix}.000195 & -.000164\\ -.000164 & .000163\end{bmatrix}\begin{bmatrix}1\\ -1.034\end{bmatrix}$$

$$\hat{g}=1.034$$
$$\hat{g}\overset{A}{\sim}N(g,0.083^2)$$

linearize

$$\frac{\beta_2}{1-\beta_3} = 1$$
$$\beta_2 = 1 - \beta_3$$
$$d = \beta_2 + \beta_3 - 1 = 0$$
$$\mathrm{var}(\beta_2 + \beta_3 - 1)$$
$$= \mathrm{var}(\beta_2) + \mathrm{var}(\beta_3) + 2\,\mathrm{cov}(\beta_2, \beta_3)$$
$$\mathrm{var}(d) = 0.000195 + 0.000163 - 2 \times .000164$$
$$= .0003$$
$$\hat{d} = .0022 \sim N(0, 0.017^2)$$

