k. Power is conditional

i. Unconditionally, is average over values of O_+

ii. Slightly lower than power evaluated at average.

- Let $\eta = \mathrm{E}\left[\sqrt{O_+}\right]$
- Suppose conditional power of form $\Phi(A + B\sqrt{O_+})$
  - ▷ $A + B\sqrt{O_+}$ is well above 0 if power well above 50%
  - ▷ Hence expect $A + B\eta > 0$
- Then unconditional power is $\mathrm{E}\left[\Phi(A + B\sqrt{O_+})\right] \approx$ $\Phi(A + B\eta) + B\phi(A + B\eta)\mathrm{E}\left[\sqrt{O_+} - \eta\right] -$ $B^2(A + B\eta)\phi(A + B\eta)\mathrm{E}\left[(\sqrt{O_+} - \eta)^2\right]/2 =$ $\Phi(A + B\eta) - B^2(A + B\eta)\phi(A + B\eta)\mathrm{Var}\left[\sqrt{O_+}\right]$
- Hence power you get putting in expectation in place of $\sqrt{O_+}$ is lower than unconditional power

iii. Hence sample size must be slightly higher

B&D2: 7.4

3. Scored exposures

a. Hypotheses

i. $H_0 : \mathrm{E}\left[O_k\right] \propto Q_k$, reflecting differences due to PYAR, but NOT exposure increase.

- Let $\pi_k^0 = Q_k/Q_+$.

ii. $H_A : \mathrm{E}[O_k] \propto \pi_k^A$, $\sum_k \pi_k^A = 1$

b. $T = \sum_{k=0}^{K-1} x_k(O_k - \pi_k^0 O_+)/O_+$

c. $\mu_0 = 0$, $\sigma_0 = \sqrt{(\sum_k x_k^2 \pi_k^0 - (\sum_k x_k \pi_k^0)^2)/O_+}$

d. $\mu_A = \sum_k x_k(\pi_k^A - \pi_k^0)$, $\sigma_A = \sqrt{(\sum_k x_k^2 \pi_k^A - (\sum_k x_k \pi_k^A)^2)/O_+}$

B&D2: 7.6

E. Two–Way Studies (2×2 tables)

1. Cohort Studies

a. Suppose

i. $\pi_0 = \mathsf{P}\,[\text{Case}|\text{Unexposed}]$

ii. $\pi_1 = \mathsf{P}\,[\text{Case}|\text{Exposed}]$

b. Hypotheses

i. $H_0 : \pi_0 = \pi_1 (= \pi_0^0)$

ii. $H_A : \psi = \pi_1(1-\pi_0)/[\pi_0(1-\pi_1)]$

- Denote alternative probabilities by π_1^A and π_0^A

iii. Typically write π_0^A and π_1^A as functions of parameter measuring distance between them and π_0^0

- In our case, function of ψ and π_0

- For instance, $\pi_0^A = \pi_0^0$ and $\pi_1^A = \frac{\pi_0 \psi}{1-\pi_0+\pi_0\psi}$

iv. Power depends on π_0

- Typically π_0^0 between π_0^A and π_1^A

c. Use standard two–sample bionomial test.

i. $\mu_0 = 0$, $\mu_A = \pi_1^A - \pi_1^0$

ii. $\sigma_0 = \sqrt{\pi_0^0(1-\pi_0^0)/O_{0+} + \pi_1^0(1-\pi_1^0)/O_{1+}}$

iii. $\sigma_A = \sqrt{\pi_0^A(1-\pi_0^A)/O_{0+} + \pi_1^A(1-\pi_1^A)/O_{1+}}$

d. Alternatively, could build test around differences in arc sine transform.

i. $\mu_0 = 0$, $\sigma_0 = \sqrt{1/(4O_{0+}) + 1/(4O_{1+})}$

ii. $\mu_A = \arcsin(\sqrt{\pi_1^A}) - \arcsin(\sqrt{\pi_0^A})$, $\sigma_A = \sigma_0$.

iii. Power $\Phi\left(-z_\alpha + \frac{\arcsin(\sqrt{\pi_1^A})-\arcsin(\sqrt{\pi_0^A})}{\sqrt{1/(4O_{0+})+1/(4O_{1+})}}\right)$

2. Case-Control Studies

a. Same as for cohort study,

i. except probabilities are for exposure rather than case

b. Suppose

i. $\rho_0 = \text{P}\left[\text{Exposed}|\text{Control}\right]$

ii. $\rho_1 = \text{P}\left[\text{Exposed}|\text{Case}\right]$

iii. $\zeta = \text{P}\left[\text{Exposed}\right]$

c. By Bayes theorem,

i. $\rho_0 = \frac{(1-\pi_1)\zeta}{(1-\pi_1)\zeta+(1-\pi_0)(1-\zeta)}$

ii. $\rho_1 = \frac{\pi_1\zeta}{\pi_1\zeta+\pi_0(1-\zeta)}$

3. Hypergeometric model, conditioning on all marginals

a. Unfortunately μ_A and σ_A do not have easy expressions.

b. Recall *Cornfield's Approximation*:

i. Solve $E_{i+} = O_{i+}$, $E_{+j} = O_{+j}$, $\frac{E_{11}E_{00}}{E_{01}E_{10}} = \psi$

ii. $E_{00} = (O_{+0} + O_{0+})/2 + \frac{O_{++}}{2(\psi-1)} - \frac{\sqrt{4\,O_{1+}\,O_{+0}\,(\psi-1)+(O_{++}+(O_{0+}-O_{+0})\,(\psi-1))^2}}{2\,(\psi-1)}$

iii. $\mathrm{Var}\,[O_{11}] \approx (\sum_{i=0}^{1}\sum_{j=0}^{1} 1/E_{ij})^{-1}$

iv. For $\psi = 1$, this is underestimate. See Figures 9 and 10.

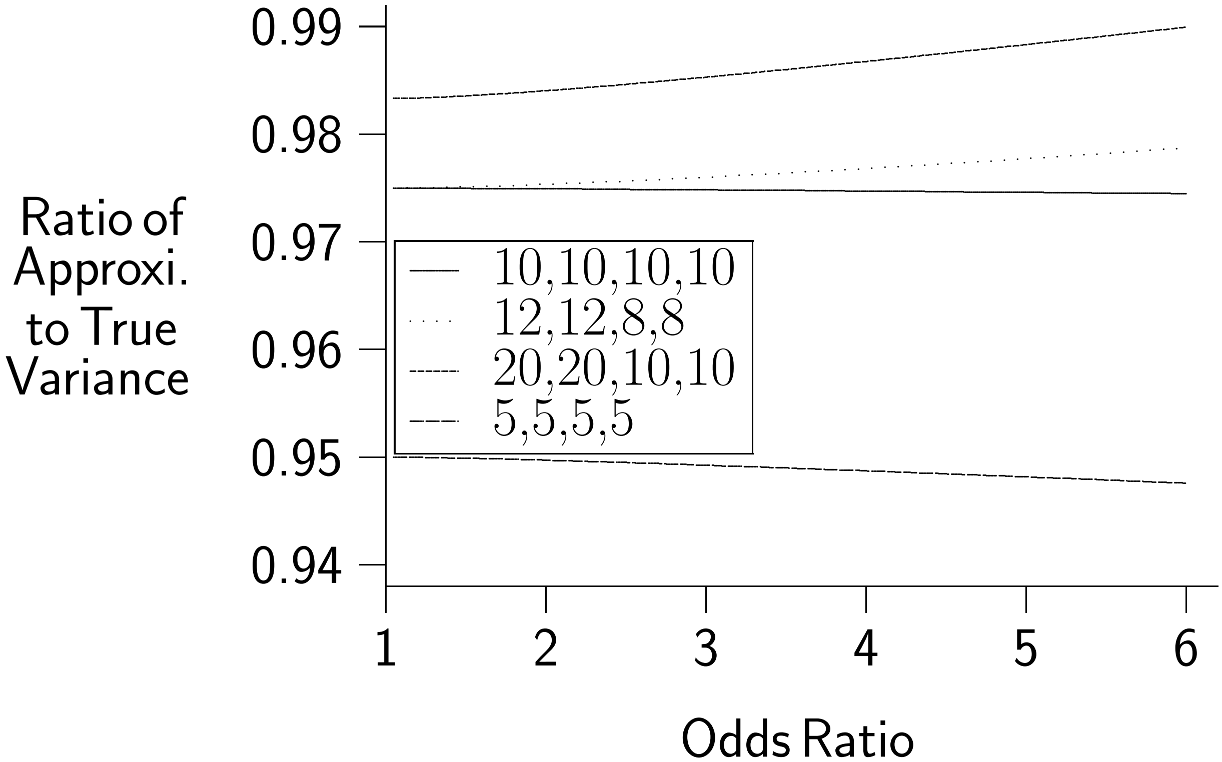
4. Example: Prostate cancer for stage 3 patients

a. Setup

i. 217 assigned high dose, 75 assigned low dose, 195 survive, 97 die

	Die	Survive	Total
Unexposed	33.82	41.18	75
Exposed	63.18	153.82	217
Total	97	195	292

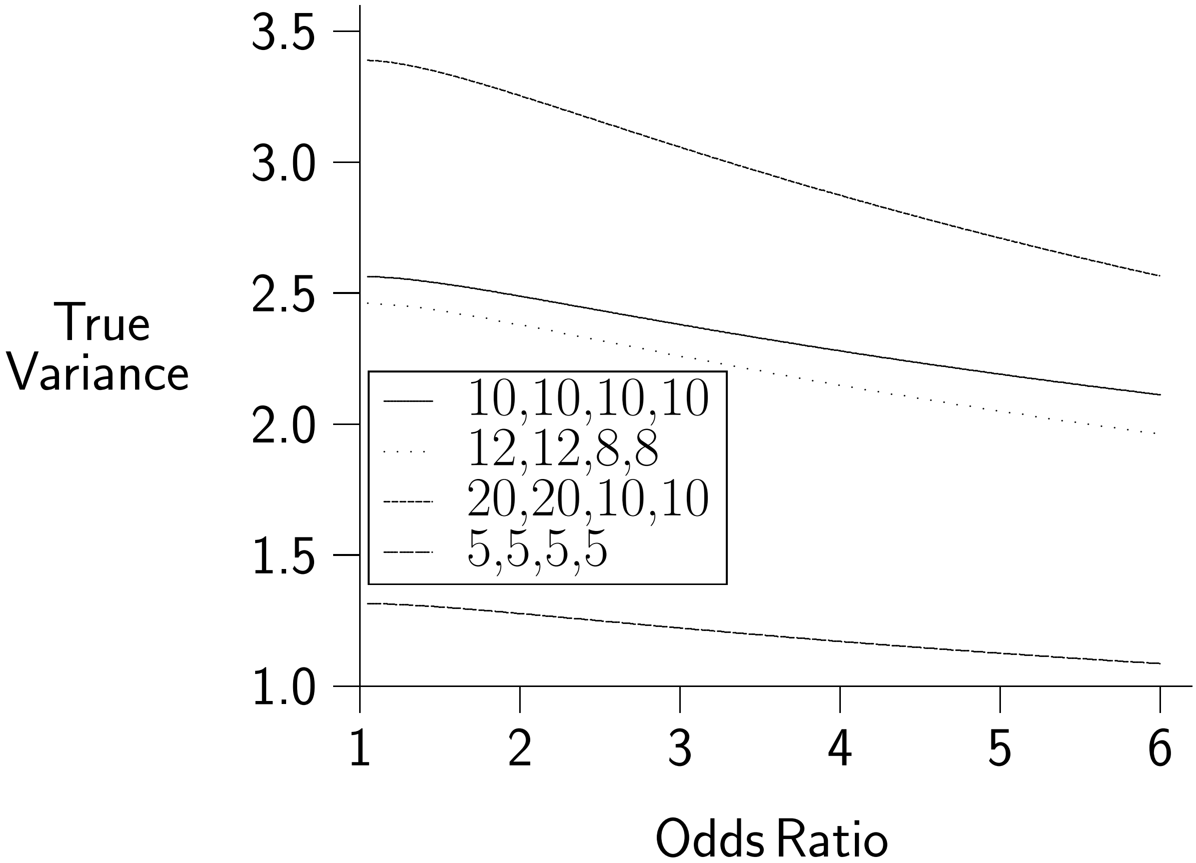
ii. Cohort Study:



- $\pi_0^0 = \pi_1^0 = 2/3$
- $\pi_0^A = 2/3\,,\ \pi_1^0 = 4/5$
- $\mu_0 = 0\,,\ \sigma_0 = \sqrt{(2/3)(1/3)/75 + (2/3)(1/3)/217} = 0.06314$
- $\mu_A = (4/5) - (2/3) = 2/15\,,\ \sigma_A = \sqrt{(2/3)(1/3)/75 + (4/5)(1/5)/217} = 0.06083$
- Power is
$$\Phi\left(\frac{2/15 - 1.96 \times 0.06314}{0.06083}\right) = 56.2\%$$

iii. Case–Control Study:

- 2 cases for every control



- Same hypotheses about π_j
- Assume proportion exposed is $217/292 \approx 3/4$.
- $H_0 : \rho_0^0 = \rho_1^0 = \frac{(1-2/3)\times(217/292)}{(1-2/3)\times(217/292)+(1-2/3)\times(75/292)} = .743$
- $H_A : \rho_0^A == \frac{(1-4/5)\times(3/4)}{(1-4/5)\times(3/4)+(1-2/3)\times(1/4)} = .643$ and $\rho_1^A = \frac{(4/5)\times(3/4)}{(4/5)\times(3/4)+(2/3)\times(1/4)} = .783$
- $\mu_0 = 0$, $\sigma_0 = \sqrt{.25 \times .75/97 + .25 \times .75/195} = 0.05380$
- $\mu_A = .783 - .643 = .140$, and $\sigma_A = \sqrt{.217 \times .783/97 + .643 \times .357/195} = 0.05441$.

- Power $\Phi\left(\frac{.140-1.96\times 0.05380}{0.05441}\right) = 73.7\%$
- To get 80% power need:
- $\tau_0 = \sqrt{.25 \times .75 + .25 \times .75/2} = 0.5303$,
  $\tau_A = \sqrt{.217 \times .783 + .643 \times .357/2} = 0.5336$.
- Need $(0.5336 \times .842 + 0.5303 \times 1.96)^2/.140^2 = 113$ controls

iv. Hypergeometric:

v. $H_0 : \psi = 1$ vs. $H_A : \pi_0^A = 2/3, \pi_1^A = 4/5 \rightarrow \psi = 2$

vi. $\mu_0 = 75\times 97/292 = 24.91$, $\sigma_0\sqrt{217 \times 75 \times 195 \times 97/(292^2 \times}$ 3.52

vii. Alt. expectation $\mu_A = (75 + 97)/2 + 292/(2 \times (2 - 1)) - \sqrt{4 \times 217 \times 97 + (292 + (75 - 97) \times (2 - 1))^2}/(2 \times (2 - 1)) = 33.82$,

$$\sigma_A = \frac{1}{\sqrt{1/33.82 + 1/41.18 + 1/63.18 + 1/153.82}} = 3.62.$$

viii. Power

$$\Phi\left(-\frac{24.91 + 1.96 \times 3.52 - 33.82}{3.62}\right) = 71\%$$

b. Conditional power is highest when $O_{1+} = O_{0+}$ (for cohort study) or $O_{+1} = O_{+0}$ (for case–control study)

i. It is dangerous to depend on this happening

ii. Accounting for variation in this margin requires consideration of the parameter eliminated by conditioning.

5. Matched designs

a. Inference will be done on discordant pairs

b. Number of discordant pairs required will be given by binomial sample size described earlier

i. Need

c. $\pi_1(1-\pi_0)+\pi_0(1-\pi_1)$ is proportion of pairs that will be discordant.

i. If matching is necessary, π_1 and π_0 will vary over strata

ii. Want average value of $\pi_1(1-\pi_0)+\pi_0(1-\pi_1)$

iii. Typically lower than same expression evaluated at average.

d. Hence divide by this to get necessary number of pairs

e. Remember that the eventual number is random.

i. Note also that distribution depends on whether alternative hypothesis is true

- If $\pi_0 > .5$, expect more discordant pairs under H_0
- If $\pi_0 < .5$, expect more discordant pairs under H_A

F. Monte Carlo Methods

1. Power calculations are often done using computer simulation.
   a. Randomly recreate data sets under null hypothesis
   b. Test hypothesis of no effect on each data sets
   c. Power is proportion of tests rejected
2. Advantages:
   a. Aren't tied to things with good normal approximation
   b. Can calculate unconditional power for conditional test easily
   c. Setting it up is more straight–forward
3. Disadvantages
   a. Need computer to do calculations
      i. If the number of trials is large, computation time might be long
   b. Results will contain some random variation

Gov: 2.3–2.5

V. Bioassay

A. Preliminary problem:

1. $Y_j = \zeta + \epsilon_j$,
2. $W_j = \mu + \delta_j$,
3. $(\epsilon_j, \delta_j) \sim \mathcal{N}\left(\mathbf{o}, \begin{pmatrix} \sigma^2 & \rho\sigma\tau \\ \rho\sigma\tau & \tau^2 \end{pmatrix}\right)$.
4. Estimate $\xi = \zeta/\mu$

5. Consider estimator $\bar{Y}/\bar{W}$

6. Problem: random variable in denomonator.

   a. Unfortunately distribution is non-standard

   b. Expectation is

$$n\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\exp(-\frac{(w-\mu)^2}{2\tau^2/n}-\frac{(y-\zeta)^2}{2\sigma^2/n})}{2\pi\tau\sigma}\frac{w}{y}\,dw\,dy$$

$$=\sqrt{n}\int_{-\infty}^{\infty}\frac{\exp(-\frac{(y-\mu)^2}{2\sigma^2/n})}{(2\pi)^{1/2}\sigma}\frac{\mu}{y}\,dy$$

      i. Integral doesn't converge absolutely.

      ii. Similar to log odds ratio case

7. Two methods for approximating th edistribution of the ratio of means

   a. Using delta method, mean and variance of approximating distribution are $\zeta/\mu = \xi$ and

$$\begin{pmatrix}\frac{1}{\mu} & -\frac{\xi}{\mu^2}\end{pmatrix}\begin{pmatrix}\sigma^2/n & \rho\sigma\tau/n\\ \rho\sigma\tau/n & \tau^2/n\end{pmatrix}\begin{pmatrix}\frac{1}{\mu}\\ -\frac{\xi}{\mu^2}\end{pmatrix} =$$

$$\mu^{-2}n^{-1}(\sigma^2-2\,\rho\,\sigma\,\tau\,\xi+\tau^2\,\xi^2)\,.$$

   b. Exact distribution:

      i. Let $U=\frac{\bar{W}-t\bar{Y}+t\mu-\zeta}{\sqrt{\tau^2/n+t^2\sigma^2/n-2t\rho\sigma\tau/n}}$ and $V=\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$

ii. Let $u = \frac{\sqrt{n}(t\mu - \zeta)}{\sqrt{\tau^2 + t^2\sigma^2 - 2t\rho\sigma\tau}}$ and $v = \frac{-\mu}{\sigma/\sqrt{n}}$.

iii. $\mathsf{P}\left[\bar{W}/\bar{Y} \le t\right]$

$$= \mathsf{P}\left[\bar{W} - t\bar{Y} \le 0 \& \bar{Y} > 0\right] + \mathsf{P}\left[\bar{W} - t\bar{Y} \ge 0 \& \bar{Y} < 0\right]$$

$$= \mathsf{P}\left[U \le u \& V > v\right] + \mathsf{P}\left[U \ge u \& V \le v\right]$$

$$= \mathsf{P}\left[U \le u\right] - \mathsf{P}\left[U \le u \& V \le v\right] + \mathsf{P}\left[U \ge u \& V \le v\right]$$

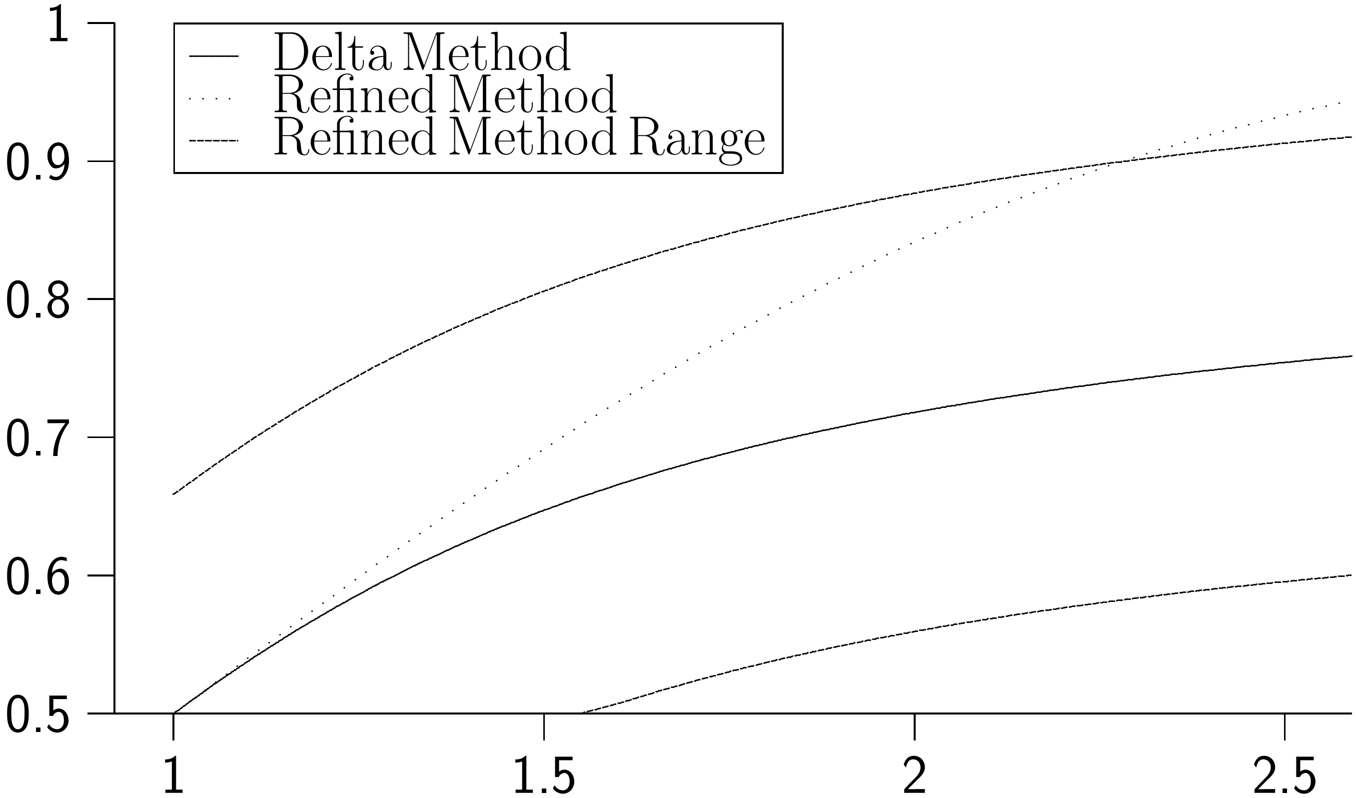
$$= \Phi\left(\frac{\sqrt{n}(t - \xi)}{\sqrt{\tau^2 + t^2\sigma^2 - 2t\rho\sigma\tau}/\mu}\right) + R$$

for $|R| \le \Phi(-\sqrt{n}\mu/\sigma)$. See Figures 11

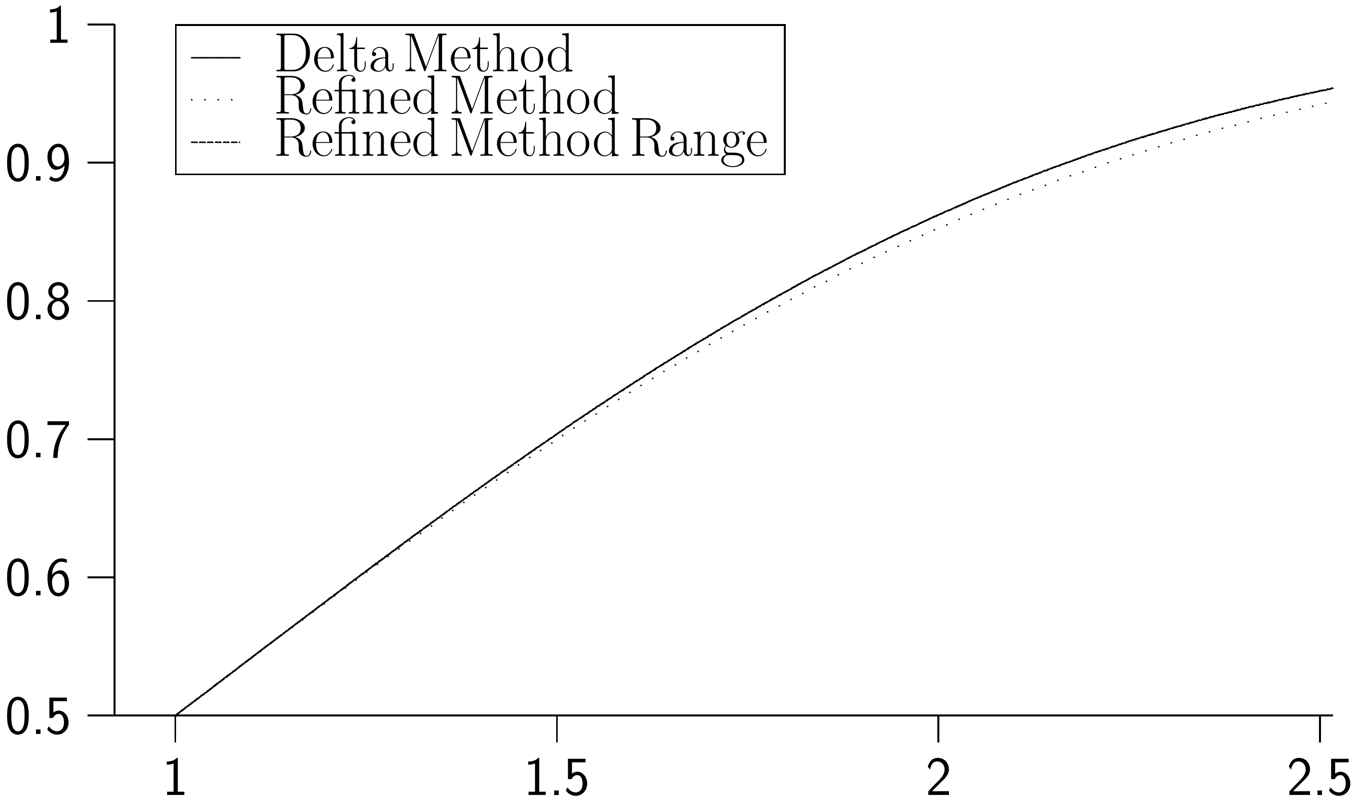
B. Real aim: confidence intervals

1. $\xi\bar{W} - \bar{Y} \sim \mathcal{N}(0, \xi^2\tau^2/n + \sigma^2/n - 2\rho\sigma\tau\xi/n)$
2. $\mathsf{P}\left[\frac{(\xi\bar{W} - \bar{Y})^2}{\xi^2\tau^2/n + \sigma^2/n - 2\rho\sigma\tau\xi/n} \le z^2_{\alpha/2}\right] = 1 - \alpha$.
3. Set of ξ satisfying statement inside probability is CI (sort of)

   a. Restriction is quadratic inequality.



$\sigma = 1\ ,\ \tau = 1\ ,\ \rho = 0.5\ ,\ \mu = 1\ ,\ \xi = 1$



$\sigma = .1\ ,\ \tau = 1\ ,\ \rho = 0.5\ ,\ \mu = 1\ ,\ \xi = 1$