Recap

- Distances between probability distributions
- Convergence in distribution
- Central limit theorem
**Theorem (Central Limit Theorem)**

Let $X_1, X_2, \ldots$, be i.i.d. from a population with an MGF that is finite near the origin. Then $X_1$ has finite mean $\mu$ and finite variance $\sigma^2$. Let

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}$$

and let $Z \sim N(0, 1)$. Then $Z_n \rightarrow Z$ in distribution, i.e.

$$P(Z_n \leq z) \rightarrow \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

for all $z$. 
Notation

- If \( (X_n - a_n)/b_n \overset{D}{\to} Z \sim N(0, 1) \), then we sometimes write
  \[
  X_n \sim AN(a_n, b_n^2)
  \]

- Read this as
  \[
  X_n \text{ is approximately normal}
  \]
  or
  \[
  X_n \text{ is asymptotically normal.}
  \]
- \( b_n^2 \) is sometimes called the *asymptotic variance*.
- \( b_n \) is sometimes called the *asymptotic standard deviation*. 
Examples

- The sample mean $\overline{X}_n$ is approximately normally distributed with mean $\mu$ and SD $\sigma/\sqrt{n}$.
- Alternatively, $\overline{X}_n \sim \text{AN}(\mu, \sigma^2/n)$.
- The sum $Y_n = \sum_{i=1}^{n} X_i$ has an approximate normal distribution with mean $n\mu$ and SD $\sqrt{n}\sigma$.
- Alternatively, $Y_n \sim \text{AN}(n\mu, n\sigma^2)$.
- If $Y_n \sim \text{Binomial}(n, p)$ then $Y_n \sim \text{AN}(np, np(1 - p))$.
  - $Y_n$ can be viewed as the sum of $n$ independent $\text{Bernoulli}(p)$ random variables.
- If $Y_n \sim \chi^2_n$ then $Y_n \sim \text{AN}(n, 2n)$.
  - $Y_n$ can be viewed as a sum of $n$ independent $\chi^2_1$ random variables.
- If $Y_\alpha \sim \text{Gamma}(\alpha, \beta)$ then $Y_\alpha \sim \text{AN}(\alpha\beta, \alpha\beta^2)$ as $\alpha \to \infty$.
- If $Y_\lambda \sim \text{Poisson}(\lambda)$ then $Y_\lambda \sim \text{AN}(\lambda, \lambda)$ as $\lambda \to \infty$. 
Proof of the Central Limit Theorem.

- Let $Y_i = (X_i - \mu)/\sigma$.
- Then $E[Y_i] = 0$, $\text{Var}(Y_i) = 1$, and $Z_n = \sqrt{n} \overline{Y}_n$.
- The MGF of $Z_n$ is therefore

$$M_{Z_n}(t) = E \left[ \exp \left\{ t \sqrt{n} \frac{1}{n} \sum Y_i \right\} \right]$$

$$= E \left[ \exp \left\{ \frac{t}{\sqrt{n}} \sum Y_i \right\} \right]$$

$$= M_{Y_1} \left( \frac{t}{\sqrt{n}} \right)^n$$

- Now by Taylor’s theorem we can write

$$M_Y(s) = M_Y(0) + sM_Y'(0) + \frac{s^2}{2} M_Y''(0) + R_Y(s)$$

where $R_Y(s) = o(s^2)$ as $s \to 0$. 
Proof of the Central Limit Theorem (continued).

• Furthermore,

\[ M_Y(0) = 1 \]
\[ M'_Y(0) = E[Y] = 0 \]
\[ M''_Y(0) = E[Y^2] = \text{Var}(Y) = 1 \]

• Therefore

\[
M_Y \left( \frac{t}{\sqrt{n}} \right)^n = \left( 1 + \frac{t^2}{2n} + R_Y \left( \frac{t}{\sqrt{n}} \right) \right)^n
\]
\[
= \left( 1 + \frac{\frac{1}{2}t^2 + nR_Y(t/\sqrt{n})}{n} \right)^n
\]
\[
\to e^{t^2/2}
\]

as \( n \to \infty \).

• This is the MGF of the standard normal distribution.
Other Limiting Distributions

Example

• Suppose $X_1, X_2, \ldots$ are non-negative i.i.d. random variables from a distribution with CDF $F$.

• Let $Y_n = \min\{X_1, \ldots, X_n\}$.

• Suppose $F(x) \sim ax^b$ for some positive $a, b$ as $x \to 0$; this means

$$\lim_{x \to 0} \frac{F(x)}{ax^b} = 1.$$  

• This implies that $F(y) > 0$ for $y > 0$.

• The CDF of $Y_n$ is $F_{Y_n}(y) = 1 - (1 - F(y))^n$.

• This converges to one for all $y > 0$, so $Y_n \overset{D}{\to} 0$ and $Y_n \overset{P}{\to} 0$.

• Some simulations for $X_i \sim \text{Uniform}[0, 1]$: http://www.stat.uiowa.edu/~luke/classes/193/convergence.R
Example (continued)

- Can we find constants $c_n$ such that $Z_n = c_n Y_n$ converges to a non-degenerate limit?
- The CDF of $Z_n$ is

$$F_{Z_n}(z) = P(Y_n \leq z/c_n) = 1 - (1 - F(z/c_n))^n = 1 - \left(1 - \frac{nF(x/c_n)}{n}\right)^n.$$  

- As $c_n \to \infty$ we have

$$nF(z/c_n) \sim na(z/c_n)^b = \frac{n}{c_n^b}az^b.$$  

- So taking $c_n = n^{1/b}$ yields $nF(z/c_n) \to az^b$ and therefore

$$F_{Z_n}(z) \to 1 - e^{-az^b}.$$  

- This is the CDF of a **Weibull** distribution.
- For a Uniform$[0, 1]$ distribution $a = 1$ and $b = 1$, so the limiting distribution is exponential with mean 1.
Example (continued)

- If $F$ is a Gamma($\alpha, 1$) CDF then $F(x) \sim \frac{x^\alpha}{\alpha \Gamma(\alpha)}$ as $x \to 0$.
- For $\alpha = 2$ this means $F(x) \sim \frac{1}{2}x^2$.
- This code provides a graphical comparison of the exact and approximate CDFs of $Z_n$ for $n = 10$:

  ```r
  n <- 10
  x <- seq(0,4,len=101)
  plot(x,1-exp(-0.5*x^2), type = "l")
  lines(x, 1 - (1 - pgamma(x/sqrt(n), 2))^n, lty = 2)
  ```
- You could also compare the true densities to the density of the approximation.
- On the original scale,

  $$F_{Y_n}(y) = F_{Z_n}(c_ny) \approx F_Z(c_ny) = 1 - e^{-a(c_ny)^b}.$$
Other Convergence Results

**Theorem (Slutsky’s Theorem)**
Suppose \( X_n \overset{D}{\to} X \) and \( Y_n \overset{P}{\to} a \), where \( a \) is a constant. Then

\[
X_n Y_n \overset{D}{\to} Xa \\
X_n + Y_n \overset{D}{\to} X + a
\]

**Theorem (Continuous Mapping Theorem)**
Suppose \( X_n \to X \) in probability (in distribution) and \( f \) has the property that

\[
P(X \in \{x : f \text{ is continuous at } x\}) = 1
\]

Then \( f(X_n) \to f(X) \) in probability (in distribution).
Example

- Suppose $X_1, X_2, \ldots$, are i.i.d with mean $\mu$ and finite variance $\sigma^2$.
- Let
  \[ \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \]
  and
  \[ S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2. \]
- We already know that
  - $\bar{X}_n \xrightarrow{P} \mu$
  - $\bar{X}_n \rightarrow \mu$ almost surely
  - $\bar{X}_n \sim \text{AN}(\mu, \sigma^2/n)$.
- What can we say about the behavior of $S_n^2$ as $n \rightarrow \infty$?
Example (continued)

- If we assume that $X_1$ has a finite fourth moment $\theta_4 = E[(X_1 - \mu)^4]$ then we have

$$E[S_n^2] = \sigma^2$$

$$\text{Var}(S_n^2) = \frac{1}{n} \left( \theta_4 - \frac{n - 3}{n - 1} \sigma^4 \right) \to 0.$$ 

- So $S_n^2 \to \sigma^2$ in mean square and in probability.
Example (continued)

- Another approach is to decompose $S_n^2$ as

\[
S_n^2 = \frac{1}{n-1} \left[ \sum_{i=1}^{n} (X_i - \mu)^2 - n(\bar{X}_n - \mu)^2 \right]
\]

\[
= \frac{n}{n-1} \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 - \frac{n}{n-1} (\bar{X}_n - \mu)^2
\]

\[
= \frac{n}{n-1} U_n - \frac{n}{n-1} V_n
\]

- By the WLLN

\[
U_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2 \xrightarrow{p} \sigma^2.
\]

- By the WLLN and the continuous mapping theorem $V_n = (\bar{X}_n - \mu)^2 \xrightarrow{p} 0$.

- Slutsky's theorem then shows that

\[
S_n^2 = \frac{n}{n-1} U_n - \frac{n}{n-1} V_n \xrightarrow{p} \sigma^2.
\]
Example (continued)

- By the SLLN we also have $U_n \to \sigma^2$ almost surely.
- By the SLLN and continuity $V_n \to 0$ almost surely as well.
- Continuity then also implies $S_n^2 \to \sigma^2$ almost surely.
Example (continued)

- Since the variance of $S_n^2$ is $O(n^{-1})$ we might expect

$$W_n = \sqrt{n}(S_n^2 - \sigma^2)$$

... to have a non-degenerate limiting distribution.

- Big picture:
  - After dropping smaller order terms $S_n^2 \approx U_n$.
  - $U_n$ is a sum of i.i.d. variables $Y_i = (X_i - \mu)^2$ with

$$E[Y_i] = E[(X_i - \mu)^2] = \sigma^2$$

$$\text{Var}(Y_i) = E[Y_i^2] - E[Y_i]^2$$

$$= E[(X_i - \mu)^4] - \sigma^4$$

$$= \theta_4 - \sigma^4.$$

- By the CLT $U_n \sim \text{AN}(\sigma^2, (\theta_4 - \sigma^4)/n)$.
- So $S_n^2 \sim \text{AN}(\sigma^2, (\theta_4 - \sigma^4)/n)$. 
Example (continued)

- To verify this carefully, write

\[ S_n^2 = U_n + \frac{1}{n-1} U_n + \frac{n}{n-1} V_n. \]

- Then

\[
\sqrt{n}(S_n^2 - \sigma^2) = \sqrt{n}(U_n - \sigma^2) + \frac{\sqrt{n}}{n-1} U_n + \frac{n}{n-1} \sqrt{n} V_n
\]

\[ = Z_n + \frac{\sqrt{n}}{n-1} U_n + \frac{n}{n-1} \sqrt{n} V_n. \]

- By the CLT \( Z_n \overset{d}{\to} Z \sim N(0, \theta^4 - \sigma^4). \)

- To complete the proof we need to show that the remaining two terms converge to zero in probability.
Example (continued)

• First term $\frac{\sqrt{n}}{n-1} U_n$:
  - $U_n \xrightarrow{P} \sigma^2$ by the WLLN, and $\frac{\sqrt{n}}{n-1} \to 0$.
  - So by Slutsky’s theorem $\frac{\sqrt{n}}{n-1} U_n \xrightarrow{P} 0$.

• Second term $\frac{n}{n-1} \sqrt{n} V_n$:
  - $\frac{n}{n-1} \to 1$ and
  - $\sqrt{n} V_n = \sqrt{n}(\bar{X}_n - \mu)^2 = [\sqrt{n}(\bar{X}_n - \mu)](\bar{X}_n - \mu)$.
    - By the CLT $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} V \sim N(0, \sigma^2)$.
    - By the WLLN $\bar{X}_n - \mu \xrightarrow{P} 0$.
    - So by Slutsky’s theorem $\frac{n}{n-1} \sqrt{n} V_n \xrightarrow{P} 0$.

• This completes the proof.
Example

• From the CLT we know that

\[ Z_n = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{D} Z \sim N(0, 1). \]

• What happens if we replace \( \sigma \) by \( S_n \) to get

\[ T_n = \frac{\bar{X}_n - \mu}{S_n / \sqrt{n}}? \]

• Write \( T_n \) as

\[ T_n = \frac{\sigma}{S_n} \left[ \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \right] = \frac{\sigma}{S_n} Z_n. \]

• By a homework problem \( \sigma / S_n \xrightarrow{P} 1. \)

• So Slutsky’s theorem implies that \( T_n \xrightarrow{D} Z. \)
Recap

- Central limit theorem
- Other Limit Distributions
- Slutsky’s theorem
- Continuous mapping theorem
Example

- What can we say about the sampling distribution of $S_n = \sqrt{S_n^2}$?
- We know that $E[S_n] < \sigma$.
- Since $\text{Var}(S_n) = E[S_n^2] - E[S_n]^2 = \sigma^2 - E[S_n]^2$ we have
  \[ E[S_n] = \sqrt{\sigma^2 - \text{Var}(S_n)}. \]
- $\text{Var}(S_n)$ is typically $O(n^{-1})$, so typically
  \[ E[S_n] = \sigma + O(n^{-1}). \]
- Since $f(x) = \sqrt{x}$ is continuous and $S_n^2 \rightarrow \sigma^2$ in probability and almost surely, we also have $S_n \rightarrow \sigma$ in probability and almost surely.
- Can we approximate the distribution of $S_n$?
- Some simulations:
  http://www.stat.uiowa.edu/~luke/classes/193/convergence.R.
Example (continued)

• We know that the distribution of $S_n^2$ is concentrated near $\sigma^2$ for large $n$.

• Near $\sigma^2$ we can approximate $f(x) = \sqrt{x}$ by its tangent

\[
f(x) \approx f(\sigma^2) + f'(\sigma^2)(x - \sigma^2)
= \sqrt{\sigma^2} + \frac{1}{2} \frac{1}{\sqrt{\sigma^2}}(x - \sigma^2)
= \sigma + \frac{1}{2\sigma}(x - \sigma^2)
\]

• So $S_n \approx \sigma + \frac{1}{2\sigma} (S_n^2 - \sigma^2)$.

• We know that $S_n^2 - \sigma^2 \sim \text{AN}(0, (\theta_4 - \sigma^4)/n)$.

• This suggests that

\[
S_n \sim \text{AN}(\sigma, (f'(\sigma^2))^2 (\theta_4 - \sigma^4)/n)
= \text{AN} \left( \sigma, \left( \frac{1}{2\sigma} \right)^2 (\theta_4 - \sigma^4)/n \right)
= \text{AN} \left( \sigma, \frac{1}{4n} \left( \frac{\theta_4}{\sigma^2} - \sigma^2 \right) \right).
\]
Example (continued)

- To verify this more carefully, Taylor’s theorem applied to the continuous, differentiable function \( f(x) = \sqrt{x} \) in a neighborhood of \( \sigma^2 \) gives

\[
f(x) = f(\sigma^2) + f'(\sigma^2)(x - \sigma^2) + o(x - \sigma^2).
\]

- Therefore

\[
\sqrt{n}(f(S_n^2) - f(\sigma^2)) = f'(\sigma^2)\sqrt{n}(S_n^2 - \sigma^2) + \sqrt{n} o(S_n^2 - \sigma^2) = U_n + V_n.
\]

- Now \( \sqrt{n}(S_n^2 - \sigma^2) \overset{D}{\to} Z \sim N(0, \theta_4 - \sigma^4) \) and \( f'(\sigma^2) \) is a constant.

- So \( U_n = f'(\sigma^2)\sqrt{n}(S_n^2 - \sigma^2) \overset{D}{\to} U \sim N(0, (f'(\sigma^2))^2(\theta_4 - \sigma^4)) \).

- The remainder is

\[
V_n = \sqrt{n} o(S_n^2 - \sigma^2) = \left[ \sqrt{n}(S_n^2 - \sigma^2) \right] \left[ \frac{o(S_n^2 - \sigma^2)}{S_n^2 - \sigma^2} \right].
\]

- This converges to zero in probability by the continuous mapping theorem and Slutsky’s theorem.
Example (continued)

- The same idea can be used to approximate the distribution of $\log S_n^2$.
- The Taylor expansion near $\sigma^2$ produces

$$\log S_n^2 \approx \log \sigma^2 + \frac{1}{\sigma^2} (S_n^2 - \sigma^2).$$

- The resulting distribution approximation is

$$\log S_n^2 \sim \text{AN} \left( \log \sigma^2, \frac{\theta_4 - \sigma^4}{n\sigma^4} \right).$$

- The log transformation
  - reduces the skewness of the distribution of $S_n^2$
  - spreads the distribution across the entire real line
- This may improve the quality of the approximation.
**Theorem (Delta Method, Propagation of Error)**

Suppose $X_n \xrightarrow{P} a$, $Y_n(X_n - a) \xrightarrow{D} Z$, and $f$ is differentiable at $a$. Then

$$Y_n(f(X_n) - f(a)) \xrightarrow{D} f'(a)Z$$

If $Y_n$ is real-valued, $X_n \in \mathbb{R}^p$, $f : \mathbb{R}^p \to \mathbb{R}^q$, then this holds with

$$f'(a) = \left( \frac{\partial f_i}{\partial x_j} \right)_{i=1,\ldots,q \atop j=1,\ldots,p}$$
• The formal Delta method is a formal version of

\[ f(X_n) \approx f(a) + f'(a)(X_n - a) \]

if \( f \) is differentiable at \( a \) and \( X_n \) is close to \( a \) with high probability.

• If \( X_n \sim AN(a, b_n^2) \) with \( b_n \to 0 \), then the delta method can be expressed as

\[
\begin{align*}
  f(X_n) &\sim AN(f(a), (|f'(a)|b_n)^2) \\
          &\sim AN(f(a), (f'(a))^2b_n^2)
\end{align*}
\]

• For \( f : \mathbb{R}^p \to \mathbb{R}^q \), if \( X_n \sim AN(a, B_n) \) with \( B_n \to 0 \), then the delta method becomes

\[
  f(X_n) \sim AN(f(a), f'(a)B_nf'(a)^T)
\]
Example

• Suppose $X_1, \ldots, X_n$ are i.i.d. positive random variables with mean $\mu$ and variance $\sigma^2$.
• Let $T_n = \log X_n$.
• Then, informally,

$$T_n \approx \log \mu + \frac{1}{\mu} (\overline{X}_n - \mu) \sim \text{AN} \left( \log \mu, \frac{1}{n \mu^2} \sigma^2 \right).$$

• More formally:
  • By the central limit theorem
    $$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{D} Z \sim \text{N}(0, \sigma^2)$$
  • Take $Y_n = \sqrt{n}$, $a = \mu$, $f(x) = \log x$ in the Delta method.
  • Then $f'(a) = 1/\mu$, and
    $$\sqrt{n}(\log \overline{X}_n - \log \mu) \xrightarrow{D} \frac{1}{\mu} Z \sim \text{N}(0, \sigma^2 / \mu^2).$$
Example (continued)

- By the continuous mapping theorem

\[
\frac{1}{\bar{X}_n} \xrightarrow{P} \frac{1}{\mu}
\]

if \( \mu \neq 0 \).

- So by Slutsky’s theorem, if \( \mu \neq 0 \),

\[
S_n/\bar{X}_n \xrightarrow{P} \sigma/\mu.
\]

- By Slutsky’s theorem therefore

\[
\sqrt{n} \frac{\bar{X}_n}{S_n} (\log \bar{X}_n - \log \mu) \xrightarrow{D} N(0, 1)
\]

- This is useful for forming approximate confidence intervals for \( \log \mu \).
Example

- Let $X \sim N(10, (0.1)^2)$ and $Y \sim N(15, (0.2)^2)$ be independent.
- $X$ and $Y$ represent measurements of the sides of a rectangle.
- An estimate of the area of the rectangle based on these measurements is $A = XY$.
- We would like an approximation to the distribution of $A$.
- The function $f(x, y) = xy$ is continuous and differentiable.
- The gradient of $f$ is

\[
\nabla f(x, y) = \left( \frac{\partial f(x, y)}{\partial x}, \frac{\partial f(x, y)}{\partial y} \right) = (y, x).
\]
Example (continued)

• The Taylor expansion of \( f \) near \((10, 15)\) is

\[
f(x, y) \approx 10 \times 15 + 15(x - 10) + 10(y - 15)
= 150 + 15(x - 10) + 10(y - 15).
\]

• So

\[
A \approx 150 + 15(X - 10) + 10(Y - 15)
\sim N(150, (15 \times 0.1)^2 + (10 \times 0.2)^2)
= N(150, (2.5)^2).
\]
Example (continued)

- The exact density of $A$ is

$$f_A(a) = \int f_X(a/y)f_Y(y)/|y| \, dy.$$ 

- This is not available in closed form but can be computed numerically.
- A graphical comparison is produced by

```r
g <- function(a) {
  f <- function(y) dnorm(a/y, 10, .1) * dnorm(y, 15, .2) / abs(y)
  integrate(f, 14, 16)$value  ## these limits seem reasonable
}
a <- seq(140, 160, len = 101)
d <- sapply(a, g)
plot(a, d, type="l")
lines(a, dnorm(a, 150, 2.5), col = "red")
```
Recap

- Delta method
- Examples of using limit theorems
Example

- A line is described by the equation
  \[ y = A + Bx. \]
- \(A\) and \(B\) are jointly normally distributed with parameters
  \[ \mu_A = 2 \quad \sigma_A = 0.2 \quad \mu_B = 1 \quad \sigma_B = 0.1 \quad \rho_{AB} = 0.5. \]
- The value of \(x\) where \(y = 0\) is \(X = -A/B\).
- We would like to approximate the distribution of \(X\).
- \(X = f(A, B)\) with \(f(a, b) = -a/b\).
- \(f\) is continuous and differentiable for \(b \neq 0\).
Example (continued)

• The gradient of \( f \) is
  \[
  \nabla f(a, b) = (-1/b, a/b^2).
  \]

• The first order Taylor approximation to \( f \) near \((a, b) = (\mu_A, \mu_B)\) is
  \[
  f(a, b) \approx -\frac{\mu_A}{\mu_B} - \frac{1}{\mu_B} (a - \mu_A) + \frac{\mu_A}{\mu_B^2} (b - \mu_B)
  = -\frac{\mu_A}{\mu_B} - \frac{1}{\mu_B} a + \frac{\mu_A}{\mu_B^2} b.
  \]
  \[
  = -2 - a + 2b.
  \]

• Therefore
  \[
  X \approx -2 - A + 2B
  \sim \mathcal{N}(-2 - \mu_A + 2\mu_B, \sigma_A^2 + 4\sigma_B^2 - 4\sigma_A\sigma_B\rho_{AB})
  = \mathcal{N}(-2, (0.2)^2).
  \]
Example (continued)

- The exact distribution is
  \[ f_X(x) = \int f_{AB}(-xb, b) |b| db. \]

- This is available in closed form but is somewhat messy.
- We can use a simple simulation to check the approximation:
  ```r
  N <- 10000
  z1 <- rnorm(N)
  z2 <- rnorm(N, 0.5 * z1, sqrt(0.75))
  A <- 2 + 0.2 * z1
  B <- 1 + 0.1 * z2
  plot(density(-A/B))
  x <- seq(-3, -1, len = 100)
  lines(x, dnorm(x, -2, 0.2), lty = 2)
  ```

- The normal approximation seems to work quite well.
- Are \( \mu_X \approx -2 \) and \( \sigma^2_X \approx (0.2)^2 \) good approximations to the mean and variance of \( X \)?
Normal Approximation to the Beta Distribution

• Suppose
  • $X_n \sim \text{Gamma}(\alpha_n, 1)$, $Y_n \sim \text{Gamma}(\beta_n, 1)$
  • $X_n, Y_n$ are independent
  • $\alpha_n \to \infty$ and $\beta_n \to \infty$.
  • $\alpha_n / (\alpha_n + \beta_n) = p$ with $0 < p < 1$.

• Then

  $X_n \sim \text{AN}(\alpha_n, \alpha_n)$
  $Y_n \sim \text{AN}(\beta_n, \beta_n)$

• A useful result:

**Theorem**

If $X_n \sim \text{AN}(a_n, b_n^2)$, $Y_n \sim \text{AN}(c_n, d_n^2)$, and $X_n, Y_n$ are independent, then

\[
\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \sim \text{AN} \left( \begin{pmatrix} a_n \\ c_n \end{pmatrix}, \begin{pmatrix} b_n^2 & 0 \\ 0 & d_n^2 \end{pmatrix} \right)
\]
For this example, this implies

\[
\begin{pmatrix}
V_n \\
W_n
\end{pmatrix} = \begin{pmatrix}
\frac{X_n}{\alpha_n + \beta_n} \\
\frac{Y_n}{\alpha_n + \beta_n}
\end{pmatrix}
\]

\[\sim \text{AN} \left( \left( \frac{\alpha_n}{\alpha_n + \beta_n}, \frac{\beta_n}{\alpha_n + \beta_n} \right), \left( \frac{\alpha_n}{(\alpha_n + \beta_n)^2}, 0, \frac{\beta_n}{(\alpha_n + \beta_n)^2} \right) \right) \]

\[\sim \text{AN} \left( \left( \frac{p}{\alpha_n + \beta_n}, \frac{1-p}{\alpha_n + \beta_n} \right), \left( 0, \frac{1-p}{\alpha_n + \beta_n} \right) \right) \]
• Now set

\[ U_n = \frac{X_n}{X_n + Y_n} = \frac{X_n}{\alpha_n + \beta_n} + \frac{Y_n}{\alpha_n + \beta_n} = \frac{V_n}{V_n + W_n} \sim \text{Beta}(\alpha_n, \beta_n) \]

• Let

\[ f(x, y) = \frac{x}{x + y} = 1 - \frac{y}{x + y} \]

• Then

\[ U_n = f(V_n, W_n). \]
• The function

\[ f(x, y) = \frac{x}{x + y} = 1 - \frac{y}{x + y} \]

is differentiable for \( x > 0, y > 0 \).

• The gradient of \( f \) is

\[ \nabla f(x, y) = \left( \frac{\partial}{\partial x} f(x, y), \frac{\partial}{\partial y} f(x, y) \right) \]

\[ = \left( \frac{y}{(x + y)^2}, -\frac{x}{(x + y)^2} \right) \]

• So

\[ \nabla f(p, 1 - p) = (1 - p, -p) \]

• Also

\[ f(p, 1 - p) = p. \]
Therefore

\[ U_n \sim \text{AN} \left( p, (1 - p, -p) \left( \frac{p}{\alpha_n + \beta_n}, 0 \right) \left( 1 - p \right) \right) \]

\[ \sim \text{AN} \left( p, \frac{p(1 - p)^2 + (1 - p)p^2}{\alpha_n + \beta_n} \right) \]

\[ \sim \text{AN} \left( p, \frac{p(1 - p)}{\alpha_n + \beta_n} \right) \]

Replacing \( p \) with its definition in terms of \( \alpha_n \) and \( \beta_n \):

\[ U_n \sim \text{AN} \left( \frac{\alpha_n}{\alpha_n + \beta_n}, \frac{\alpha_n \beta_n}{(\alpha_n + \beta_n)^3} \right) \]
• Comparing exact and approximate densities:

```r
x <- seq(0, 1, len = 100)
n <- 10
p <- 0.25
plot(x, dbeta(x, p * n, (1 - p) * n), type = "l")
lines(x, dnorm(x, p, sqrt(p * (1 - p) / n)), lty = 2)
```

• Standardized to make approximation N(0,1):

```r
z <- seq(-3, 3, len = 100)
s <- sqrt(p * (1 - p) / n)
plot(z, s * dbeta(p + z * s, p * n, (1 - p) * n), type = "l")
lines(z, dnorm(z), lty = 2)
```
• Comparing standardized CDFs:

```r
plot(z, pbeta(p + z * s, p * n, (1 - p) * n), type = "l")
lines(z, pnorm(z), lty = 2)
```

• A numerical summary (Kolmogorov distance):

```r
max(abs(pbeta(p + z * s, p * n, (1 - p) * n) - pnorm(z)))
```

• Summary for several sample sizes:

```r
nn <- c(10, 20, 40, 60, 100, 150, 200)
pp <- sapply(nn, function(n) {
  s <- sqrt(p * (1 - p) / n)
  max(abs(pbeta(p + z * s, p * n, (1 - p) * n) - pnorm(z)))
})
plot(nn, pp)
```
• A probability plot:

\[ F_1 \leftarrow \text{pbeta}(p + z \ast s, p \ast n, (1 - p) \ast n) \]
\[ F_2 \leftarrow \text{pnorm}(z) \]
\[ \text{plot}(F_1, F_2, \text{type} = "l") \]
\[ \text{abline}(0, 1, \text{lty} = 2) \]

• A Quantile plot:

\[ \text{plot}(p + s \ast z, \text{qbeta}(\text{pnorm}(z), p \ast n, (1 - p) \ast n), \text{type} = "l") \]
\[ \text{abline}(0, 1, \text{lty} = 2) \]

• Relative errors:

\[ \text{plot}(z, (F_1 - F_2) / (F_1 + F_2), \text{type} = "l") \]
\[ \text{lines}(z, (F_1 - F_2) / (1 - F_1 + 1 - F_2), \text{lty} = 2) \]

• A numerical summary for relative errors:

\[ \text{max}(\text{abs}(F_1 - F_2) / (F_1 + F_2), \text{abs}(F_1 - F_2) / (1 - F_1 + 1 - F_2)) \]

• A better approximation in the tails is obtained by approximating the distribution of

\[ \log(U/(1 - U)). \]
Approximate Distribution of Order Statistics

• For $0 < p < 1$ the $p$-th sample quantile for a $U[0, 1]$ population satisfies
  \[ U(\{np\}) \sim \text{Beta}(\{np\}, n - \{np\} + 1). \]

• The normal approximation for the Beta distribution therefore implies that
  \[ U(\{np\}) \sim \text{AN}(p, p(1 - p)/n). \]

• This means that
  \[ U(\{np\}) \xrightarrow{p} p \]

• and
  \[ \sqrt{n}(U(\{np\}) - p) \xrightarrow{D} V \sim N(0, p(1 - p)) \]
• Suppose $F$ is a population CDF such that
  • $F$ is a continuous distribution with density $f$
  • At the $p$-th population quantile $F^{-1}(p)$

  $$F'(F^{-1}(p)) = f(F^{-1}(p)) > 0.$$ 

• Since

  $$X_{(k)} \sim F^{-1}(U_{(k)})$$

we can approximate the distribution of $X_{\{np\}}$ using
  • the normal approximation to the distribution of $U_{\{np\}}$
  • the delta method
The normal approximation to the distribution of $U_{\{np\}}$ is

$$U_{\{np\}} \sim \text{AN}(p, p(1 - p)/n).$$

The delta method then shows that

$$X_{\{np\}} = F^{-1}(U_{\{np\}}) \sim \text{AN}\left(F^{-1}(p), \left(\frac{d}{dp} F^{-1}(p)\right)^2 \frac{p(1 - p)}{n}\right)$$

Now

$$\frac{d}{dp} F^{-1}(p) = \frac{1}{f(F^{-1}(p))}$$

So

$$X_{\{np\}} \sim \text{AN}\left(F^{-1}(p), \frac{p(1 - p)}{n f(F^{-1}(p))^2}\right)$$
Example

- Suppose \( f \sim N(\mu, \sigma^2) \) and \( p = 1/2 \). Then \( F^{-1}(1/2) = \mu \) and

\[
f(F^{-1}(p)) = \frac{1}{\sqrt{2\pi}\sigma}
\]

- Then

\[
\tilde{X} \sim AN \left( \mu, \frac{2\pi\sigma^2}{4n} \right)
\]

- Now \( \bar{X} \sim N(\mu, \sigma^2/n) \) and

\[
\sqrt{\frac{2\pi}{4}} \approx 1.2533 > 1
\]

- So for a given sample size \( \bar{X} \) is a more accurate estimator of \( \mu \) than \( \tilde{X} \).
Higher Order Delta Method

• The first order delta method works well when the derivative of the function is not zero.
• For example, suppose $X_n \sim \text{Binomial}(n, p)$.
• Then $Y_n = X_n/n \sim \text{AN}(p, p(1-p)/n)$.
• Let $S_n = Y_n(1-Y_n)$.
• The first order delta method will produce a useful approximate distribution for $0 < p < 1$ and $p \neq 1/2$.
• For $p = 1/2$ a second order Taylor expansion of $f(p) = p(1-p)$ can be used.
• This is called the second order delta method.
• Other approaches, for example based on simulation, may be more effective.