Large Sample Tools*

STA 312: Fall 2012

Background Reading: Davison's *Statistical models*

- For completeness, look at Section 2.1, which presents some basic applied statistics in an advanced way.
- Especially see Section 2.2 (Pages 28-37) on convergence.
- Section 3.3 (Pages 77-90) goes more deeply into simulation than we will. At least skim it.

Overview

Contents

*See last slide for copyright information.

1 Foundations

Sample Space Ω, $\omega \in \Omega$

- Observe whether a single individual is male or female: $\Omega = \{F, M\}$
- Pair of individuals; observe their genders in order: $\Omega = \{(F,F),(F,M),(M,F),(M,M)\}$
- Select n people and count the number of females: $\Omega = \{0,\ldots,n\}$

For limits problems, the points in Ω are infinite sequences.

Random variables are functions from Ω into the set of real numbers

$$Pr\{X \in B\} = Pr(\{\omega \in \Omega : X(\omega) \in B\})$$

Random Sample $X_1(\omega),\ldots,X_n(\omega)$

- $T = T(X_1,\ldots,X_n)$
- $T = T_n(\omega)$
- Let $n \to \infty$ to see what happens for large samples

Modes of Convergence

- Almost Sure Convergence
- Convergence in Probability
- Convergence in Distribution

Almost Sure Convergence

We say that T_n converges *almost surely* to T, and write $T_n \stackrel{a.s.}{\to}$ if

$$Pr\{\omega : \lim_{n\to\infty} T_n(\omega) = T(\omega)\} = 1.$$

- Acts like an ordinary limit, except possibly on a set of probability zero.
- All the usual rules apply.
- Called convergence with probability one or sometimes strong convergence.

2 Law of Large Numbers

Strong Law of Large Numbers

Let $X_1, \ldots, X_n$ be independent with common expected value μ.

$$\overline{X}_n \stackrel{a.s.}{\rightarrow} E(X_i) = \mu$$

The only condition required for this to hold is the existence of the expected value.

Probability is long run relative frequency

- Statistical experiment: Probability of "success" is p
- Carry out the experiment many times independently.
- Code the results $X_i = 1$ if success, $X_i = 0$ for failure, $i = 1, 2, \ldots$

Sample proportion of successes converges to the probability of success Recall $X_i = 0$ or 1.

$$\begin{aligned} E(X_i) &= \sum_{x=0}^{1} x \, Pr\{X_i = x\} \\ &= 0 \cdot (1-p) + 1 \cdot p \\ &= p \end{aligned}$$

Relative frequency is

$$\frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}_n \stackrel{a.s.}{\rightarrow} p$$

Simulation

- Estimate almost any probability that's hard to figure out
- Power
- Weather model
- Performance of statistical methods
- Confidence intervals for the estimate

A hard elementary problem

- Roll a fair die 13 times and observe the number each time.
- What is the probability that the sum of the 13 numbers is divisible by 3?

Simulate from a multinomial

```
> nsim = 1000 # nsim is the Monte Carlo sample size
> set.seed(9999) # So I can reproduce the numbers if desired.
> kount = numeric(nsim)
> for(i in 1:nsim)
+     {
+     tot = sum(rmultinom(1,13,die)*(1:6))
+     kount[i] = (tot/3 == floor(tot/3))
+     # Logical will be converted to numeric
+     }

> kount[1:20]
 [1] 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0

> xbar = mean(kount); xbar
[1] 0.329
```

Check if the sum is divisible by 3

```
> tot = sum(rmultinom(1,13,die)*(1:6))
> tot
[1] 42
> tot/3 == floor(tot/3)
[1] TRUE
> 42/3
[1] 14
```

Estimated Probability

```
> nsim = 1000 # nsim is the Monte Carlo sample size
> set.seed(9999) # So I can reproduce the numbers if desired.
> kount = numeric(nsim)
> for(i in 1:nsim)
+     {
+     tot = sum(rmultinom(1,13,die)*(1:6))
+     kount[i] = (tot/3 == floor(tot/3))
+     # Logical will be converted to numeric
+     }

> kount[1:20]
 [1] 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0

> xbar = mean(kount); xbar
[1] 0.329
```

Confidence Interval $\overline{X} \pm z_{\alpha/2}\sqrt{\frac{\overline{X}(1-\overline{X})}{n}}$

```
> z = qnorm(0.995); z
[1] 2.575829
> pnorm(z)-pnorm(-z) # Just to check
[1] 0.99

> margerror99 = sqrt(xbar*(1-xbar)/nsim)*z; margerror99
[1] 0.03827157

> cat("Estimated probability is ",xbar," with 99% margin of error ",
+      margerror99,"\n")

Estimated probability is  0.329  with 99% margin of error  0.03827157

> cat("99% Confidence interval from ",xbar-margerror99," to ",
+      xbar+margerror99,"\n")

99% Confidence interval from  0.2907284  to  0.3672716
```

Recall the Change of Variables formula: Let $Y = g(X)$

$$E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$$

Or, for discrete random variables

$$E(Y) = \sum_y y\, p_{_Y}(y) = \sum_x g(x)\, p_{_X}(x)$$

This is actually a big theorem, not a definition.

Applying the change of variables formula To approximate $E[g(X)]$

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n g(X_i) &= \frac{1}{n}\sum_{i=1}^n Y_i \stackrel{a.s.}{\rightarrow} E(Y) \\ &= E(g(X)) \end{aligned}$$

So for example

$$\frac{1}{n}\sum_{i=1}^n X_i^k \stackrel{a.s.}{\rightarrow} E(X^k)$$

$$\frac{1}{n}\sum_{i=1}^n U_i^2 V_i W_i^3 \stackrel{a.s.}{\rightarrow} E(U^2 V W^3)$$

That is, sample moments converge almost surely to population moments.

Approximate an integral: $\int_{-\infty}^{\infty} h(x)\,dx$ **Where** $h(x)$ **is a nasty function.**
Let $f(x)$ be a density with $f(x) > 0$ wherever $h(x) \neq 0$.

$$\begin{aligned} \int_{-\infty}^{\infty} h(x)\,dx &= \int_{-\infty}^{\infty} \frac{h(x)}{f(x)} f(x)\,dx \\ &= E\left[\frac{h(X)}{f(X)}\right] \\ &= E[g(X)], \end{aligned}$$

So

- Sample $X_1, \ldots, X_n$ from the distribution with density $f(x)$
- Calculate $Y_i = g(X_i) = \frac{h(X_i)}{f(X_i)}$ for $i = 1, \ldots, n$
- Calculate $\overline{Y}_n \stackrel{a.s.}{\to} E[Y] = E[g(X)]$

Convergence in Probability

We say that T_n converges *in probability* to T, and write $T_n \stackrel{P}{\to} T$ if for all $\epsilon > 0$,

$$\lim_{n\to\infty} P\{|T_n - T| < \epsilon\} = 1$$

Convergence in probability (say to a constant θ) means no matter how small the interval around θ, for large enough n (that is, for all $n > N_1$) the probability of getting that close to θ is as close to one as you like.

Weak Law of Large Numbers

$$\overline{X}_n \stackrel{p}{\to} \mu$$

- Almost Sure Convergence implies Convergence in Probability
- Strong Law of Large Numbers implies Weak Law of Large Numbers

3 Consistency

Consistency $T = T(X_1, \ldots, X_n)$ **is a statistic estimating a parameter** θ

The statistic T_n is said to be *consistent* for θ if $T_n \stackrel{P}{\to} \theta$.

$$\lim_{n\to\infty} P\{|T_n - \theta| < \epsilon\} = 1$$

The statistic T_n is said to be *strongly consistent* for θ if $T_n \stackrel{a.s.}{\to} \theta$.
Strong consistency implies ordinary consistency.

Consistency is great but it's not enough.

- It means that as the sample size becomes indefinitely large, you probably get as close as you like to the truth.
- It's the least we can ask. Estimators that are not consistent are completely unacceptable for most purposes.

$$T_n \stackrel{a.s.}{\to} \theta \Rightarrow U_n = T_n + \frac{100,000,000}{n} \stackrel{a.s.}{\to} \theta$$

Consistency of the Sample Variance

$$\begin{aligned} \widehat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 \\ &= \frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2 \end{aligned}$$

By SLLN, $\overline{X}_n \stackrel{a.s.}{\to} \mu$ and $\frac{1}{n}\sum_{i=1}^n X_i^2 \stackrel{a.s.}{\to} E(X^2) = \sigma^2 + \mu^2$.
Because the function $g(x,y) = x - y^2$ is continuous,

$$\widehat{\sigma}_n^2 = g\left(\frac{1}{n}\sum_{i=1}^n X_i^2, \overline{X}_n\right) \stackrel{a.s.}{\to} g(\sigma^2+\mu^2, \mu) = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

4 Central Limit Theorem

Convergence in Distribution Sometimes called *Weak Convergence*, or *Convergence in Law*

Denote the cumulative distribution functions of $T_1, T_2, \ldots$ by $F_1(t), F_2(t), \ldots$ respectively, and denote the cumulative distribution function of T by $F(t)$.

We say that T_n converges *in distribution* to T, and write $T_n \stackrel{d}{\to} T$ if for every point t at which F is continuous,

$$\lim_{n\to\infty} F_n(t) = F(t)$$

Univariate Central Limit Theorem

Let $X_1, \ldots, X_n$ be a random sample from a distribution with expected value μ and variance σ^2. Then

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \stackrel{d}{\rightarrow} Z \sim N(0,1)$$

Connections among the Modes of Convergence

- $T_n \stackrel{a.s.}{\rightarrow} T \Rightarrow T_n \stackrel{p}{\rightarrow} T \Rightarrow T_n \stackrel{d}{\rightarrow} T$.
- If a is a constant, $T_n \stackrel{d}{\rightarrow} a \Rightarrow T_n \stackrel{p}{\rightarrow} a$.

Sometimes we say the distribution of the sample mean is approximately normal, or asymptotically normal.

- This is justified by the Central Limit Theorem.
- But it does not mean that $\overline{X}_n$ converges in distribution to a normal random variable.
- The Law of Large Numbers says that $\overline{X}_n$ converges in distribution to a constant, μ.
- So $\overline{X}_n$ converges to μ in distribution as well.

Why would we say that for large n, the sample mean is approximately $N(\mu, \frac{\sigma^2}{n})$?

Have $Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \stackrel{d}{\rightarrow} Z \sim N(0,1)$.

$$\begin{aligned}
Pr\{\overline{X}_n \leq x\} &= Pr\left\{\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq \frac{\sqrt{n}(x-\mu)}{\sigma}\right\} \\
&= Pr\left\{Z_n \leq \frac{\sqrt{n}(x-\mu)}{\sigma}\right\} \\
&\approx \Phi\left(\frac{\sqrt{n}(x-\mu)}{\sigma}\right)
\end{aligned}$$

Suppose Y is *exactly* $N(\mu, \frac{\sigma^2}{n})$:

$$\begin{aligned}
Pr\{Y \leq x\} &= Pr\left\{\frac{\sqrt{n}(Y - \mu)}{\sigma} \leq \frac{\sqrt{n}(x-\mu)}{\sigma}\right\} \\
&= Pr\left\{Z_n \leq \frac{\sqrt{n}(x-\mu)}{\sigma}\right\} \\
&= \Phi\left(\frac{\sqrt{n}(x-\mu)}{\sigma}\right)
\end{aligned}$$

5 Convergence of random vectors

Convergence of random vectors

1. Definitions (All quantities in boldface are vectors in $\mathbb{R}^m$ unless otherwise stated)

   - $\star$ $\mathbf{T}_n \stackrel{a.s.}{\to} \mathbf{T}$ means $P\{\omega : \lim_{n\to\infty} \mathbf{T}_n(\omega) = \mathbf{T}(\omega)\} = 1$.
   - $\star$ $\mathbf{T}_n \stackrel{P}{\to} \mathbf{T}$ means $\forall\epsilon > 0$, $\lim_{n\to\infty} P\{||\mathbf{T}_n - \mathbf{T}|| < \epsilon\} = 1$.
   - $\star$ $\mathbf{T}_n \stackrel{d}{\to} \mathbf{T}$ means for every continuity point $\mathbf{t}$ of $F_{\mathbf{T}}$, $\lim_{n\to\infty} F_{\mathbf{T}_n}(\mathbf{t}) = F_{\mathbf{T}}(\mathbf{t})$.

2. $\mathbf{T}_n \stackrel{a.s.}{\to} \mathbf{T} \Rightarrow \mathbf{T}_n \stackrel{P}{\to} \mathbf{T} \Rightarrow \mathbf{T}_n \stackrel{d}{\to} \mathbf{T}$.

3. If $\mathbf{a}$ is a vector of constants, $\mathbf{T}_n \stackrel{d}{\to} \mathbf{a} \Rightarrow \mathbf{T}_n \stackrel{P}{\to} \mathbf{a}$.

4. Strong Law of Large Numbers (SLLN): Let $\mathbf{X}_1, \ldots \mathbf{X}_n$ be independent and identically distributed random vectors with finite first moment, and let $\mathbf{X}$ be a general random vector from the same distribution. Then $\overline{\mathbf{X}}_n \stackrel{a.s.}{\to} E(\mathbf{X})$.

5. Central Limit Theorem: Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. random vectors with expected value vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then $\sqrt{n}(\overline{\mathbf{X}}_n - \boldsymbol{\mu})$ converges in distribution to a multivariate normal with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$.

6. Slutsky Theorems for Convergence in Distribution:

   (a) If $\mathbf{T}_n \in \mathbb{R}^m$, $\mathbf{T}_n \stackrel{d}{\to} \mathbf{T}$ and if $f : \mathbb{R}^m \to \mathbb{R}^q$ (where $q \leq m$) is continuous except possibly on a set C with $P(\mathbf{T} \in C) = 0$, then $f(\mathbf{T}_n) \stackrel{d}{\to} f(\mathbf{T})$.

   (b) If $\mathbf{T}_n \stackrel{d}{\to} \mathbf{T}$ and $(\mathbf{T}_n - \mathbf{Y}_n) \stackrel{P}{\to} 0$, then $\mathbf{Y}_n \stackrel{d}{\to} \mathbf{T}$.

   (c) If $\mathbf{T}_n \in \mathbb{R}^d$, $\mathbf{Y}_n \in \mathbb{R}^k$, $\mathbf{T}_n \stackrel{d}{\to} \mathbf{T}$ and $\mathbf{Y}_n \stackrel{P}{\to} \mathbf{c}$, then

   $$\begin{pmatrix} \mathbf{T}_n \\ \mathbf{Y}_n \end{pmatrix} \stackrel{d}{\to} \begin{pmatrix} \mathbf{T} \\ \mathbf{c} \end{pmatrix}$$

7. Slutsky Theorems for Convergence in Probability:

   (a) If $\mathbf{T}_n \in \mathbb{R}^m$, $\mathbf{T}_n \stackrel{P}{\to} \mathbf{T}$ and if $f : \mathbb{R}^m \to \mathbb{R}^q$ (where $q \leq m$) is continuous except possibly on a set C with $P(\mathbf{T} \in C) = 0$, then $f(\mathbf{T}_n) \stackrel{P}{\to} f(\mathbf{T})$.

   (b) If $\mathbf{T}_n \stackrel{P}{\to} \mathbf{T}$ and $(\mathbf{T}_n - \mathbf{Y}_n) \stackrel{P}{\to} 0$, then $\mathbf{Y}_n \stackrel{P}{\to} \mathbf{T}$.

   (c) If $\mathbf{T}_n \in \mathbb{R}^d$, $\mathbf{Y}_n \in \mathbb{R}^k$, $\mathbf{T}_n \stackrel{P}{\to} \mathbf{T}$ and $\mathbf{Y}_n \stackrel{P}{\to} \mathbf{Y}$, then

   $$\begin{pmatrix} \mathbf{T}_n \\ \mathbf{Y}_n \end{pmatrix} \stackrel{P}{\to} \begin{pmatrix} \mathbf{T} \\ \mathbf{Y} \end{pmatrix}$$

8. Delta Method (Theorem of Cramér, Ferguson p. 45): Let $g : \mathbb{R}^d \to \mathbb{R}^k$ be such that the elements of $\dot{\mathrm{g}}(\mathbf{x}) = \left[\frac{\partial g_i}{\partial x_j}\right]_{k\times d}$ are continuous in a neighborhood of $\boldsymbol{\theta} \in \mathbb{R}^d$. If $\mathbf{T}_n$ is a sequence of d-dimensional random vectors such that $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \stackrel{d}{\to} \mathbf{T}$, then $\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \stackrel{d}{\to} \dot{\mathrm{g}}(\boldsymbol{\theta})\mathbf{T}$. In particular, if $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \stackrel{d}{\to} \mathbf{T} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, then $\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \stackrel{d}{\to} \mathbf{Y} \sim N(\mathbf{0}, \dot{\mathrm{g}}(\boldsymbol{\theta})\boldsymbol{\Sigma}\dot{\mathrm{g}}(\boldsymbol{\theta})')$.

An application of the Slutsky Theorems

- Let $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} ?(\mu, \sigma^2)$
- By CLT, $Y_n = \sqrt{n}(\overline{X}_n - \mu) \stackrel{d}{\rightarrow} Y \sim N(0, \sigma^2)$
- Let $\widehat{\sigma}_n$ be *any* consistent estimator of σ.
- Then by 6.6c, $\mathbf{T}_n = \begin{pmatrix} Y_n \\ \widehat{\sigma}_n \end{pmatrix} \stackrel{d}{\rightarrow} \begin{pmatrix} Y \\ \sigma \end{pmatrix} = \mathbf{T}$
- The function $f(x, y) = x/y$ is continuous except if $y = 0$ so by 6.6a,

$$f(\mathbf{T}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\widehat{\sigma}_n} \stackrel{d}{\rightarrow} f(\mathbf{T}) = \frac{Y}{\sigma} \sim N(0, 1)$$

6 Delta Method

Univariate delta method

In the multivariate Delta Method 8, the matrix $\dot{\mathrm{g}}(\boldsymbol{\theta})$ is a Jacobian. The univariate version of the delta method says

$$\sqrt{n}\,(g(T_n) - g(\theta)) \stackrel{d}{\rightarrow} g'(\theta)\,T.$$

If $T \sim N(0, \sigma^2)$, it says

$$\sqrt{n}\,(g(T_n) - g(\theta)) \stackrel{d}{\rightarrow} Y \sim N\left(0, g'(\theta)^2\,\sigma^2\right).$$

A variance-stabilizing transformation An application of the delta method

- Because the Poisson process is such a good model, count data often have approximate Poisson distributions.
- Let $X_1, \ldots, X_n \stackrel{i.i.d}{\sim}$ Poisson(λ)
- $E(X_i) = Var(X_i) = \lambda$
- $Z_n = \frac{\sqrt{n}(\overline{X}_n - \lambda)}{\sqrt{\overline{X}_n}} \stackrel{d}{\rightarrow} Z \sim N(0, 1)$
- An approximate large-sample confidence interval for λ is

$$\overline{X}_n \pm z_{\alpha/2}\sqrt{\frac{\overline{X}_n}{n}}$$

- Can we do better?

Variance-stabilizing transformation continued

- CLT says $\sqrt{n}(\overline{X}_n - \lambda) \stackrel{d}{\to} T \sim N(0, \lambda)$.
- Delta method says $\sqrt{n}\left(g(\overline{X}_n) - g(\lambda)\right) \stackrel{d}{\to} g'(\lambda)\, T = Y \sim N\left(0, g'(\lambda)^2\, \lambda\right)$
- If $g'(\lambda) = \frac{1}{\sqrt{\lambda}}$, then $Y \sim N(0, 1)$.

An elementary differential equation: $g'(x) = \frac{1}{\sqrt{x}}$ **Solve by separation of variables**

$$
\begin{aligned}
& \frac{dg}{dx} = x^{-1/2} \\
\Rightarrow\ & dg = x^{-1/2}\, dx \\
\Rightarrow\ & \int dg = \int x^{-1/2}\, dx \\
\Rightarrow\ & g(x) = \frac{x^{1/2}}{1/2} + c = 2x^{1/2} + c
\end{aligned}
$$

We have found

$$
\begin{aligned}
\sqrt{n}\left(g(\overline{X}_n) - g(\lambda)\right) &= \sqrt{n}\left(2\overline{X}_n^{1/2} - 2\lambda^{1/2}\right) \\
&\stackrel{d}{\to} Z \sim N(0, 1)
\end{aligned}
$$

So,

- We could say that $\sqrt{\overline{X}_n}$ is asymptotically normal, with (asymptotic) mean $\sqrt{\lambda}$ and (asymptotic) variance $\frac{1}{4n}$.
- This calculation could justify a square root transformation for count data.
- How about a better confidence interval for λ?

Seeking a better confidence interval for λ

$$\begin{aligned}
1-\alpha &\approx \Pr\{-z_{\alpha/2} < Z < z_{\alpha/2}\} \\
&= \Pr\{-z_{\alpha/2} < 2\sqrt{n}\left(\overline{X}_n^{1/2} - \lambda^{1/2}\right) < z_{\alpha/2}\} \\
&= \Pr\left\{\sqrt{\overline{X}_n} - \frac{z_{\alpha/2}}{2\sqrt{n}} < \sqrt{\lambda} < \sqrt{\overline{X}_n} + \frac{z_{\alpha/2}}{2\sqrt{n}}\right\} \\
&= \Pr\left\{\left(\sqrt{\overline{X}_n} - \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2 < \lambda < \left(\sqrt{\overline{X}_n} + \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2\right\},
\end{aligned}$$

where the last equality is valid provided $\sqrt{\overline{X}_n} - \frac{z_{\alpha/2}}{2\sqrt{n}} \geq 0$.

Compare the confidence intervals

Variance-stabilized CI is

$$\begin{aligned}
&\left(\left(\sqrt{\overline{X}_n} - \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2 \;,\; \left(\sqrt{\overline{X}_n} + \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2\right) \\
=\;& \left(\overline{X}_n - 2\sqrt{\overline{X}_n}\frac{z_{\alpha/2}}{2\sqrt{n}} + \frac{z_{\alpha/2}^2}{4n} \;,\; \overline{X}_n + 2\sqrt{\overline{X}_n}\frac{z_{\alpha/2}}{2\sqrt{n}} + \frac{z_{\alpha/2}^2}{4n}\right) \\
=\;& \left(\overline{X}_n - z_{\alpha/2}\sqrt{\frac{\overline{X}_n}{n}} + \frac{z_{\alpha/2}^2}{4n} \;,\; \overline{X}_n + z_{\alpha/2}\sqrt{\frac{\overline{X}_n}{n}} + \frac{z_{\alpha/2}^2}{4n}\right)
\end{aligned}$$

Compare to the ordinary (Wald) CI

$$\left(\overline{X}_n - z_{\alpha/2}\sqrt{\frac{\overline{X}_n}{n}} \;,\; \overline{X}_n + z_{\alpha/2}\sqrt{\frac{\overline{X}_n}{n}}\right)$$

Variance-stabilized CI is just like the ordinary CI

Except shifted to the right by $\frac{z_{\alpha/2}^2}{4n}$.

- If there is a difference in performance, we will see it for small n.
- Try some simulations.
- Is the coverage probability closer?

Try $n = 10$, True $\lambda = 1$ Illustrate the code first

```
> # Variance stabilized Poisson CI
> n = 10; lambda=1; m=10; alpha = 0.05; set.seed(9999)
> z = qnorm(1-alpha/2)
> cover1 = cover2 = NULL
> for(sim in 1:m)
+     {
+     x = rpois(n,lambda); xbar = mean(x); xbar
+     a1 = xbar - z*sqrt(xbar/n); b1 = xbar + z*sqrt(xbar/n)
+     shift = z^2/(4*n)
+     a2 = a1+shift; b2 = b1+shift
+     cover1 = c(cover1,(a1 < lambda && lambda < b1))
+     cover2 = c(cover2,(a2 < lambda && lambda < b2))
+     } # Next sim
> rbind(cover1,cover2)
       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
cover1 TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE FALSE
cover2 TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE FALSE
> mean(cover1)
[1] 0.9
```

Code for Monte Carlo sample size = 10,000 simulations

```
# Now the real simulation
n = 10; lambda=1; m=10000; alpha = 0.05; set.seed(9999)
z = qnorm(1-alpha/2)
cover1 = cover2 = NULL
for(sim in 1:m)
    {
    x = rpois(n,lambda); xbar = mean(x); xbar
    a1 = xbar - z*sqrt(xbar/n); b1 = xbar + z*sqrt(xbar/n)
    shift = z^2/(4*n)
    a2 = a1+shift; b2 = b1+shift
    cover1 = c(cover1,(a1 < lambda && lambda < b1))
    cover2 = c(cover2,(a2 < lambda && lambda < b2))
    } # Next sim
p1 = mean(cover1); p2 = mean(cover2)
# 99 percent margins of error
me1 = qnorm(0.995)*sqrt(p1*(1-p1)/m); me1 = round(me1,3)
me2 = qnorm(0.995)*sqrt(p1*(1-p1)/m); me2 = round(me2,3)
cat("Coverage of ordinary CI = ",p1,"plus or minus ",me1,"\n")
cat("Coverage of variance-stabilized CI = ",p2,
"plus or minus ",me2,"\n")
```

Results for $n = 10$, $\lambda = 1$ and 10,000 simulations

```
Coverage of ordinary CI =  0.9292 plus or minus  0.007

Coverage of variance-stabilized CI =  0.9556 plus or minus  0.007
```

```
> p2-me2
[1] 0.9486
```

Results for $n = 100$ $\lambda = 1$ and 10,000 simulations

```
Coverage of ordinary CI =  0.9448 plus or minus  0.006

Coverage of variance-stabilized CI =  0.9473 plus or minus  0.006

> p1+me1
[1] 0.9508
```

The arcsin-square root transformation For proportions

Sometimes, variable values consist of proportions, one for each case.

- For example, cases could be hospitals.
- The variable of interest is the proportion of patients who came down with something *unrelated* to their reason for admission – hospital-acquired infection.
- This is an example of *aggregated data*.

The advice you often get

When a proportion is the response variable in a regression, use the *arcsin square root* transformation.

That is, if the proportions are $P_1, \ldots, P_n$, let

$$Y_i = \sin^{-1}(\sqrt{P_i})$$

and use the Y_i values in your regression.

Why?

It's a variance-stabilizing transformation.

- The proportions are little sample means: $P_i = \frac{1}{m}\sum_{j=1}^m X_{i,j}$
- Drop the i for now.
- $X_1, \ldots, X_m$ may not be independent, but let's pretend.
- $P = \overline{X}_m$
- Approximately, $\overline{X}_m \sim N\left(\theta, \frac{\theta(1-\theta)}{m}\right)$
- Normality is good.
- Variance that depends on the mean θ is not so good.

Apply the delta method

Central Limit Theorem says

$$\sqrt{m}(\overline{X}_m - \theta) \xrightarrow{d} T \sim N\left(0, \theta(1-\theta)\right)$$

Delta method says

$$\sqrt{m}\left(g(\overline{X}_m) - g(\theta)\right) \xrightarrow{d} Y \sim N\left(0, g'(\theta)^2\,\theta(1-\theta)\right).$$

Want a function $g(x)$ with

$$g'(x) = \frac{1}{\sqrt{x(1-x)}}$$

Try $g(x) = \sin^{-1}\left(\sqrt{x}\right)$.

Chain rule to get $\frac{d}{dx}\sin^{-1}\left(\sqrt{x}\right)$

"Recall" that $\frac{d}{dx}\sin^{-1}(x) = \frac{1}{\sqrt{1-x^2}}$. Then,

$$\begin{aligned}
\frac{d}{dx}\sin^{-1}\left(\sqrt{x}\right) &= \frac{1}{\sqrt{1-\sqrt{x}^2}} \cdot \frac{1}{2}x^{-1/2} \\
&= \frac{1}{2\sqrt{x(1-x)}}.
\end{aligned}$$

Conclusion:

$$\sqrt{m}\left(\sin^{-1}\left(\sqrt{\overline{X}_m}\right) - \sin^{-1}\left(\sqrt{\theta}\right)\right) \xrightarrow{d} Y \sim N\left(0, \frac{1}{4}\right)$$

So the arcsin-square root transformation stabilizes the variance

- The variance no longer depends on the probability that the proportion is estimating.
- Does not quite *standardize* the proportion, but that's okay for regression.
- Potentially useful for non-aggregated data too.

- If we want to do a regression on aggregated data, the point we have reached is that approximately,

$$Y_i \sim N\left(\sin^{-1}\left(\sqrt{\theta_i}\right), \frac{1}{4m_i}\right)$$

That was fun, but it was all univariate.

Because

- The multivariate CLT establishes convergence to a multivariate normal, and
- Vectors of MLEs are approximately multivariate normal for large samples, and
- The multivariate delta method can yield the asymptotic distribution of useful functions of the MLE vector,

We need to look at random vectors and the multivariate normal distribution.