Large sample tools[1]

STA442/2101 Fall 2014

[1] See last slide for copyright information.

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

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 $T_n \stackrel{a.s.}{\rightarrow} T$ if

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

- 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.

Strong Law of Large Numbers

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

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

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 θ
- 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.

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

Relative frequency is

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

Simulation

Using pseudo-random number generation by computer

- Estimate almost any probability that's hard to figure out
- Power
- Weather model
- Performance of statistical methods
- Need confidence intervals for estimated probabilities.

Estimating power by simulation

Recall the one versus two-sample t test example (chimney vent damper)

- With paired data and a positive correlation, we suspected that the two-sample test would have diminished power.
- Maybe wrong Type I error probability, too.
- Power of the correct test can be obtained analytically – more later.
- Power and Type I error probability of the *wrong* test can be more challenging.

Strategy for estimating power by simulation

Similar approach for probability of Type I error

- Generate a large number of random data sets under the alternative hypothesis.
- For each data set, test H_0.
- Estimated power is the proportion of times H_0 is rejected.

Power of the t-tests by Simulation

An example

- (X_i, Y_i) bivariate normal
- Equal Variances: $\sigma_1^2 = \sigma_2^2 = \sigma^2 = 1$
- $|\mu_1 - \mu_2| = \frac{\sigma}{2}$, so let $\mu_1 = 1, \mu_2 = 1.5$
- $Corr(X_i, Y_i) = +0.50$
- $n = 25$
- What is the power of the correct test and the incorrect test?

Simulate From a Multivariate Normal

```
rmvn <- function(nn,mu,sigma)
# Returns an nn by kk matrix, rows are independent
# MVN(mu,sigma)
    {
    kk <- length(mu)
    dsig <- dim(sigma)
    if(dsig[1] != dsig[2]) stop("Sigma must be square.")
    if(dsig[1] != kk)
        stop("Sizes of sigma and mu are inconsistent.")
    ev <- eigen(sigma,symmetric=T)
    sqrl <- diag(sqrt(ev$values))
    PP <- ev$vectors
    ZZ <- rnorm(nn*kk) ; dim(ZZ) <- c(kk,nn)
    rmvn <- t(PP%*%sqrl%*%ZZ+mu)
    rmvn
    }# End of function rmvn
```

Simulation Code

```
set.seed(9999)
n = 25; r = 0.5; nsim=1000
crit1 = qt(0.975,n-1); crit2 = qt(0.975,2*(n-1))
Mu = c(1,1.5); Sigma = rbind(c(1,r),
                             c(r,1))
nsig1 = nsig2 = 0
for(sim in 1:nsim)
    {
    dat = rmvn(n,Mu,Sigma); X = dat[,1]; Y = dat[,2]
    sig1 = t.test(x=X,y=Y,paired=T)$p.value<0.05
    if(sig1) nsig1=nsig1+1
    sig2 = t.test(x=X,y=Y,var.equal=T)$p.value<0.05
    if(sig2) nsig2=nsig2+1
    }
cat(" \n")
cat(" Based on ",nsim," simulations, Estimated Power \n")
cat(" Matched t-test: ",round(nsig1/nsim,3),"\n")
cat(" Two-sample t-test: ",round(nsig2/nsim,3),"\n")
cat(" \n")
```

Output

```
 Based on  1000  simulations, Estimated Power
 Matched t-test:  0.675
 Two-sample t-test:  0.385

Mu = c(1,1) # H0 is true -- estimate significance level

 Based on  1000  simulations, Estimated Power
 Matched t-test:  0.063
 T-sample t-test:  0.006

Based on  10000  simulations, Estimated Power
 Matched t-test:  0.053
 Two-sample t-test:  0.007
```

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 $E[g(X)]$

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

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 $\overline{Y}_n \stackrel{a.s.}{\rightarrow} E[Y] = E[g(X)]$
- Confidence interval for $\mu = E[g(X)]$ is routine.

Convergence in Probability

We say that T_n converges *in probability* to T, and write $T_n \stackrel{P}{\rightarrow} T$

if for all $\epsilon > 0$,

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

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

$$\overline{X}_n \xrightarrow{p} \mu$$

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

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 $T_n \stackrel{P}{\rightarrow} \theta$.

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

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

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,

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

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 $T_n \stackrel{d}{\rightarrow} T$ 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

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

Connections among the Modes of Convergence

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

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 almost surely (and in probability) 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* $N(\mu, \frac{\sigma^2}{n})$:

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

Convergence of random vectors I

1. Definitions (All quantities in boldface are vectors in $\mathbb{R}^m$ unless otherwise stated)
   - $\mathbf{T}_n \stackrel{a.s.}{\rightarrow} \mathbf{T}$ means $P\{\omega : \lim_{n\to\infty} \mathbf{T}_n(\omega) = \mathbf{T}(\omega)\} = 1$.
   - $\mathbf{T}_n \stackrel{P}{\rightarrow} \mathbf{T}$ means $\forall \epsilon > 0$, $\lim_{n\to\infty} P\{||\mathbf{T}_n - \mathbf{T}|| < \epsilon\} = 1$.
   - $\mathbf{T}_n \stackrel{d}{\rightarrow} \mathbf{T}$ 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. $\mathbf{T}_n \stackrel{a.s.}{\rightarrow} \mathbf{T} \Rightarrow \mathbf{T}_n \stackrel{P}{\rightarrow} \mathbf{T} \Rightarrow \mathbf{T}_n \stackrel{d}{\rightarrow} \mathbf{T}$.
3. If $\mathbf{a}$ is a vector of constants, $\mathbf{T}_n \stackrel{d}{\rightarrow} \mathbf{a} \Rightarrow \mathbf{T}_n \stackrel{P}{\rightarrow} \mathbf{a}$.
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 $\overline{\mathbf{X}}_n \stackrel{a.s.}{\rightarrow} E(\mathbf{X})$.
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 $\boldsymbol{\Sigma}$.

Convergence of random vectors II

6. Slutsky Theorems for Convergence in Distribution:
   1. If $\mathbf{T}_n \in \mathbb{R}^m$, $\mathbf{T}_n \xrightarrow{d} \mathbf{T}$ 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 $f(\mathbf{T}_n) \xrightarrow{d} f(\mathbf{T})$.
   2. If $\mathbf{T}_n \xrightarrow{d} \mathbf{T}$ and $(\mathbf{T}_n - \mathbf{Y}_n) \xrightarrow{P} 0$, then $\mathbf{Y}_n \xrightarrow{d} \mathbf{T}$.
   3. If $\mathbf{T}_n \in \mathbb{R}^d$, $\mathbf{Y}_n \in \mathbb{R}^k$, $\mathbf{T}_n \xrightarrow{d} \mathbf{T}$ and $\mathbf{Y}_n \xrightarrow{P} \mathbf{c}$, then

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

An application of the Slutsky Theorems

- Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} ?(\mu, \sigma^2)$
- By CLT, $Y_n = \sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} Y \sim N(0, \sigma^2)$
- Let $\widehat{\sigma}_n$ be *any* consistent estimator of σ.
- Then by 6.3, $\mathbf{T}_n = \begin{pmatrix} Y_n \\ \widehat{\sigma}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} Y \\ \sigma \end{pmatrix} = \mathbf{T}$
- The function $f(x, y) = x/y$ is continuous except if $y = 0$ so by 6.1,

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

We need more tools

Because

- The multivariate CLT establishes convergence to a multivariate normal, and
- Vectors of MLEs are approximately multivariate normal for large samples, and
- Most real-life models have multiple parameters,

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

Copyright Information