MTH 452 Mathematical Statistics

Instructor: Orlando Merino

University of Rhode Island

Spring Semester, 2006

5.1 Introduction

An Experiment: In 10 consecutive trips to the free throw line, a professional basketball player makes the first 6 baskets and misses the next 4 baskets.

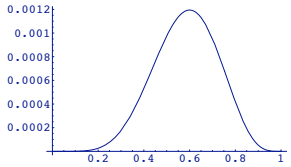
If p = probability of a making a basket,
what is a reasonable value for p?

ANSWER 1:

$$p = \frac{\#\text{ of successes}}{\#\text{ of trials}} = \frac{6}{10}$$

ANSWER 2:

The event "SSSSSSFFFF" has probability
$f(p) = p^6(1-p)^4$



To find where the maximium occurs, take derivative:

$$f'(p) = 6p^5(1-p)^4 + p^6 \cdot 4(1-p)^3 = 2p^5(1-p)^3(5p-3)$$

Therefore, $f'(p) = 0 \Rightarrow p = 1$ or $p = 0$ or $p = 3/5$

Conclusion: Select $p = 3/5$, which maximizes $f(p)$ in $[0,1]$.

What is Statistical Inference?

Data generated in accordance with certain unknown probability distribution must be analyzed and some *inference* about the unknown distribution has to be made.

In some problems, the probability distribution which generated the experimental data is completely known except for one or more *parameters*, and the problem is to make inferences about the values of the unknown parameters.

For example, suppose it is known that the distribution of heights in a population of individuals is normal with some mean μ and variance σ^2. By observing the height data for a random sample of individuals we may make inferences about μ and σ^2.

In our discussion we shall use $\theta_1, \theta_2, \ldots$ to denote parameters. The set Ω of all values of the parameters is called parameter space.

Common Distributions

Distribution	Probability Fn	Variable Range
Hypergeometric	$\frac{\binom{r}{k}\cdot\binom{w}{n-k}}{\binom{r+w}{n}}$	$k = 0, 1, \ldots, n$
Binomial(n,p)	$\binom{n}{k}p^k(1-p)^{n-k}$	$k = 0, 1, \ldots, n$
Poisson(λ)	$\frac{e^{-\lambda}\lambda^k}{k!}$	$k = 0, 1, 2, \ldots$
Geometric(p)	$p(1-p)^{k-1}$	$k = 1, 2, \ldots$
Neg.Bin.(r,p)	$\binom{k-1}{r-1}p^r(1-p)^{k-r}$	$k = r, r+1, \ldots$
Uniform on (a,b)	$\frac{1}{b-a}$	$a < y < b$
Normal(μ,σ)	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-\mu)^2}{2\sigma^2}}$	$-\infty < y < \infty$
Exponential(λ)	$\lambda e^{-\lambda y}$	$0 \le y < \infty$
Gamma(r,λ)	$\frac{\lambda^r}{(r-1)!}y^{r-1}e^{-\lambda y}$	$0 \le y$

5.2 Part 1: Maximum Likelihood

GIVEN:

- $W_1, \ldots, W_n$ a random sample from continuous pdf $f_Y(y;\theta)$ with unknown parameter θ.
- data values $W_1 = \omega_1, \ldots, W_n = \omega_n$

DEFINE: the Likelihood Function

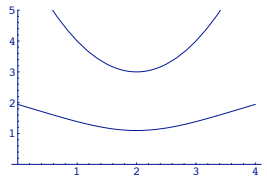
$$L(\theta) := \prod_{\ell=1}^{n} f_W(\omega_\ell, \theta) = f_W(\omega_1, \theta)\cdots f_W(\omega_n, \theta)$$

---

the MAXIMUM LIKELIHOOD ESTIMATOR for θ (or MLE) is a number $\hat{\theta}$ such that $L(\hat{\theta}) \ge L(\theta)$, for all θ

An observation about optimization that will make our life easier later

IF $f(x)$ is a positive function, then both $f(x)$ and $\ln(f(x))$ attain local minima (or maxima) at the same locations x.



In other words:

to find x that minimizes (or maximizes) $f(x)$,
it is enough to
find x that minimizes (or maximizes) $\ln(f(x))$.

Example 5.2.1 Suppose $k_1, \ldots, k_n$ are n observations of a RV X with pdf $f_X(k) = (1-p)^{k-1}p$, $k = 1, 2, 3, \ldots n$.
Find the MLE for p=probability of success.

ANSWER: The likelihood function is

$$L(p) = (1-p)^{k_1-1}p \cdots (1-p)^{k_n-1}p = (1-p)^{k-n} \cdot p^n$$

where for convenience we set $k = k_1 + \cdots + k_n$.

Take the logarithm of $L(p)$:

$$\ln L(p) = (k-n)\ln(1-p) + n\ln p$$

The derivative of "$\ln L(p)$" is

$$\frac{d}{dp}(\ln L(p)) = (k-n) \cdot \frac{-1}{1-p} + \frac{n}{p}$$

The critical points of $\ln L(p)$ are found by setting the derivative to zero:

$$(k-n) \cdot \frac{-1}{1-p} + \frac{n}{p} = 0 \;\Rightarrow\; p(n-k) + n(1-p) = 0$$

Solve for p above to get $\hat{p}$:

$$\hat{p} = \frac{n}{k}$$

Example 5.2.2 Five data points

$$Y_1 = 9.2,\; Y_2 = 5.6,\; Y_3 = 18.4,\; Y_4 = 12.1,\; Y_5 = 10.7$$

were taken from the pdf

$$f_Y(y;\theta) = \frac{1}{\theta^2}\, y\, e^{-y/\theta}, \quad 0 < y < \infty;\; 0 < \theta < \infty$$

Find a reasonable estimate for θ.

ANSWER:

$$L(\theta) \;=\; \tfrac{1}{\theta^2}\, y_1\, e^{-y_1/\theta} \;\cdots\; \tfrac{1}{\theta^2}\, y_5\, e^{-y_5/\theta}$$

$$=\; \theta^{-10}\, y_1 \cdots y_5\, e^{-(1/\theta)y}$$

where $y = y_1 + \cdots + y_5$. Hence,

$$\ln L(\theta) = -10\ln\theta + \ln(y_1 \cdots y_5) - \frac{1}{\theta}y$$

Set the derivative equal to zero:

$$\frac{d}{dp}(\ln L(p)) = \frac{-10}{\theta} + \frac{1}{\theta^2} \cdot y = 0$$

to get

$$\hat{\theta} = \frac{1}{10}y = \frac{1}{10}(y_1 + \cdots + y_5) = \frac{56.0}{10} = 5.6$$

Example 5.2.3: MLE when derivatives fail
Suppose $y_1, \ldots, y_n$ are measurements representing the pdf

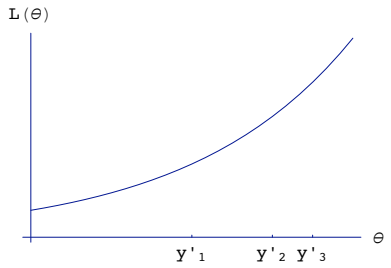
$$f_Y(y;\theta) = e^{-(y-\theta)}, \quad \theta \le y; \quad \theta > 0$$

Find the MLE for θ.

ANSWER:

$$L(\theta) = \prod_{\ell=1}^{n} e^{-(y_\ell - \theta)} = e^{-(\sum_{\ell=1}^{n} y_\ell - n\theta)} = Ce^{n\theta}$$

The following is a plot of the likelihood function:



The largest possible value of θ maximizes the likelihood function. But it is required that $\theta \le y_1$, and $\theta \le y_2$, ..., and $\theta \le y_n$. Then

$$\theta \le y_1' := \min\{y_1, \ldots, y_n\} \;\Rightarrow\; \hat{\theta} = y_1'$$

Finding MLEs with 2 or more parameters

If the model depends on 2 parameters θ_1 and θ_2 (what is an example?), finding MLEs requires solving the system of equations

$$\begin{cases} \dfrac{\partial L(\theta_1,\theta_2)}{\partial \theta_1} &= 0 \\ \dfrac{\partial L(\theta_1,\theta_2)}{\partial \theta_2} &= 0 \end{cases}$$

In general, MLEs for

k parameter models

require the solution of a system of

k equations and k unknowns

, which are a direct generalization of the system given above.

Example 5.2.4 A random sample of size n is drawn from a two parameter normal pdf. Use the method of maximum likelihood to find formulas for $\hat{\mu}$ and $\widehat{\sigma^2}$

ANSWER: (To be given in class)

5.2 Part 2: The Method of Moments

Suppose Y is a continuous RV with pdf $f_Y(y;\theta_1,\ldots,\theta_n)$. The k-th moment of Y is

$$E(Y^k) = \int_{-\infty}^{\infty} y^k f_Y(y;\theta_1,\ldots,\theta_n)dy, \quad k = 1,2,3,\ldots$$

Given a random sample $(Y_1,\ldots,Y_n)$, the corresponding **k-th sample moment** is

$$\frac{1}{n}\sum_{\ell=1}^{n} y_\ell^k = \frac{1}{n}\left(y_1^k + y_2^k + \cdots + y_n^k\right)$$

The sample moment is an approximation to the moment. We may form a system of equations, which has to be solved for $\theta_1, \theta_2, \ldots$:

$$\begin{cases} \displaystyle\int_{-\infty}^{\infty} y^1 f_Y(y;\theta_1,\ldots,\theta_n)dy &= \dfrac{1}{n}\displaystyle\sum_{\ell=1}^{n} y_\ell^1 \\ \vdots \quad & \vdots \\ \displaystyle\int_{-\infty}^{\infty} y^k f_Y(y;\theta_1,\ldots,\theta_n)dy &= \dfrac{1}{n}\displaystyle\sum_{\ell=1}^{n} y_\ell^k \end{cases}$$

Example 5.2.5 Given the random sample

$$Y_1 = 0.42,\ Y_2 = 0.10,\ Y_3 = 0.65,\ Y_4 = 0.23,$$

drawn from the pdf

$$f_Y(y;\theta) = \theta\, y^{\theta-1}, \quad 0 \le y \le 1$$

Find the method of moments estimate for θ.

ANSWER: The first moment of Y is

$$E(Y) = \int_0^1 y\theta y^{\theta-1}\,dy = \int_0^1 \theta y^{\theta}\,dy$$

$$= \frac{\theta}{\theta+1} y^{\theta+1}\Big|_0^1 = \frac{\theta}{\theta+1}$$

then

$$\frac{\theta}{\theta+1} = \frac{1}{n}\sum_{\ell=1}^{n} y_\ell = \frac{1}{4}(0.42+0.10+0.65+0.23) = 0.35$$

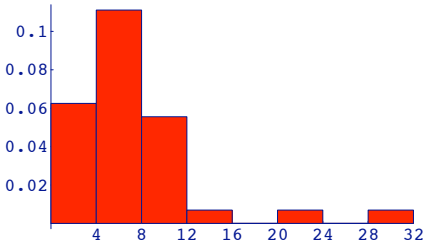
Solving for θ we get the estimate

$$\hat{\theta} = \frac{0.35}{1-0.35} = 0.54$$

Case Study 5.2.2 The following is the maximum 24-hour precipitation (in inches) for 36 inland hurricanes (1900-1969) over the Appalachians, as recorded by the U.S. Weather Bureau:

31.00, 2.82, 3.98, 4.02, 9.50, 4.50, 11.40, 10.71, 6.31, 4.95, 5.64, 5.51, 13.40, 9.72, 6.47, 10.16, 4.21, 11.60, 4.75, 6.85, 6.25, 3.42, 11.80, 0.80, 3.69, 3.10, 22.22, 7.43, 5.00, 4.58, 4.46, 8.00, 3.73, 3.50, 6.20, 0.67

Here is a (density scaled) histogram of the data



The histogram's profile suggests that Y, the maximum 24-hour precipitation can be modeled by the two-parameter gamma pdf,

$$f_Y(y;r,\lambda) = \frac{\lambda^r}{\Gamma(r)}\, y^{r-1}\, e^{-\lambda y}$$

Knowing that $E(Y) = r/\lambda$ and $Var(Y) = r/\lambda^2$, estimate λ and r using the method of moments, then plot $f_Y(y;r,\lambda)$ together with the histogram.

ANSWER: Since $Var(Y) = E(Y^2) - (E(Y))^2$,

$$E(Y^2) = \frac{r}{\lambda^2} + \frac{r^2}{\lambda^2} = \frac{r(r+1)}{\lambda^2}$$

From the data, the sample moments are

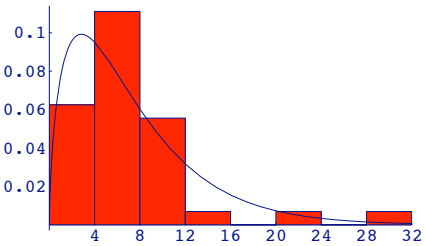
$$\frac{1}{36}\sum_{\ell=1}^{36} y_\ell = 7.2875 \quad \text{and} \quad \frac{1}{36}\sum_{\ell=1}^{36} y_\ell^2 = 85.5894$$

The two equations to solve are:

$$\frac{r}{\lambda} = 7.2875 \quad \text{and} \quad \frac{r(r+1)}{\lambda^2} = 85.5894$$

From the first equation we get $r = 7.2875\lambda$, which when substituted into the 2nd equation gives

$$\frac{7.2875\lambda(7.2875\lambda+1)}{\lambda^2} = 85.5894 \Rightarrow \begin{cases} \lambda = 0.2243 \\ r = 1.635 \end{cases}$$



$$f_Y = \frac{0.09673\, y^{0.635}}{e^{0.2243 y}}$$

5.3 Interval Estimation

The problem with point estimates:

THEY GIVE NO INDICATION ABOUT PRECISION

A way to deal with this problem is to construct a confidence interval.

A confidence interval is an interval of numbers that has "high probability" of containing the unknown parameter as an interior point.

The width of the confidence interval gives a sense of the estimator's precision.

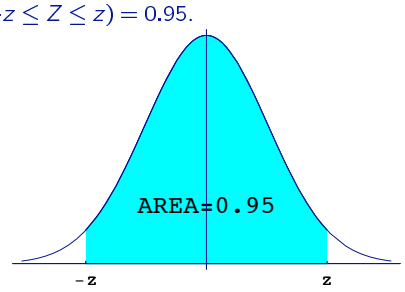
Review of $\overline{Y}$: If $Y_1, Y_2, \ldots, Y_n$ is a random sample (that is, independent, identically distributed RVs), define

$$\overline{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n)$$

Fact: If Φ= dist'n w/mean μ, std.dev. σ,

$$Y_\ell \sim \Phi(\mu, \sigma) \text{ for } 1 \le \ell \le n \;\Rightarrow\; \overline{Y} \sim \Phi(\mu, \sigma/\sqrt{n})$$

A Useful Problem: If $Z \sim N(0,1)$ (standard normal), find z so that $P(-z \le Z \le z) = 0.95$.



ANSWER: note that $P(Z \le z) = 0.975$.
The standard normal table gives $z = 1.96$.

Example 5.3.1:
Construction of a 95% Confidence Interval
A sample $Y_1 = 6.5, Y_2 = 9.2, Y_3 = 9.9, Y_4 = 12.4$ is taken from a normal pdf with $\sigma = 0.8$ and unknown μ.

We know the following facts:

- **(1)** The MLE is $\hat{\mu} = \overline{Y} = \frac{1}{n}\sum_\ell^n Y_\ell$,
  in our case, $= \frac{1}{4}(38.0)$
- **(2)** $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, the standard normal dist.
- **(3)** $P(-1.96 \le Z \le 1.96) = 0.95$

Combine (1) and (2) and (3) to get

$$P\left(-1.96 \le \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \le 1.96\right) = P(-1.96 \le Z \le 1.96) = 0.95$$

Then

$$P\left(-1.96\frac{0.8}{\sqrt{4}} \le \overline{Y} - \mu \le 1.96\frac{0.8}{\sqrt{4}}\right) = 0.95$$

Equivalently,

$$P\left(-1.96\tfrac{0.8}{\sqrt{4}} \le \mu - \overline{Y} \le 1.96\tfrac{0.8}{\sqrt{4}}\right)$$

$$= P\left(\overline{Y} - 1.96\tfrac{0.8}{\sqrt{4}} \le \mu \le \overline{Y} + 1.96\tfrac{0.8}{\sqrt{4}}\right)$$

$$= 0.95$$

That is,

$$P\left(9.5 - 1.96\tfrac{0.8}{\sqrt{4}} \le \mu \le 9.5 + 1.96\tfrac{0.8}{\sqrt{4}}\right)$$

$$= P(8.72 \le \mu \le 10.28)$$

$$= 0.95$$

Recall μ is a constant.
The 95% confidence interval (8.72,10.28)
has a 95% chance of containing μ.

More precisely,

if many intervals are computed from samples using this procedure, approximately 95% of them will contain μ.

Case Study 5.3.1 The sizes of 84 Etruscan skulls from archeological digs have sample mean $\bar{y} = 143.8$ mm. Skill widths of present day italians have a mean of 132.4 mm and a standard deviation of 6.0 mm. What can be said about the statement that the italians and etruscans share the same ethnic origins?

ANSWER: Construct a 95% confidence interval for the true mean of the population of etruscans, and determine if 132.4 lies in it. If not, it may be argued that italians and etruscans aren't related.

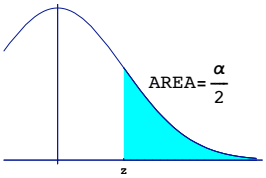
The endpoints of a 95% confidence interval for μ are given by
$\left(\bar{y} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}\ ,\ \bar{y} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) =$
$\left(143.8 - 1.96 \cdot \frac{6.0}{\sqrt{84}}\ ,\ 143.8 - 1.96 \cdot \frac{6.0}{\sqrt{84}}\right) = (142.5, 145.1)$
Conclusion: since $132.4 \notin (142.5, 145.1)$, a sample mean of 143.8 based on a sample of size 84 is not likely to come from a population where $\mu = 132.4$.

Confidence intervals in general

A $100(1-\alpha)\%$ confidence interval for μ is obtained as follows:

First find $z_{\frac{\alpha}{2}}$ defined by the equation

$$P(Z \geq z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$$



In practice $z_{\frac{\alpha}{2}}$ is found with table/computer.

The $100(1-\alpha)\%$ confidence interval for μ is

$$\left(\bar{y} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \ , \ \bar{y} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Confidence Interval for binomial parameter p

The following is a Theorem we don't prove here:

IF $X \sim$ binomial(n,p) witn n largeTHEN $\frac{X/n-p}{\sqrt{\frac{(X/n)(1-X/n)}{n}}} \approx N(0,1)$

From this relation we have

$$P\left(-z_{\frac{\alpha}{2}} \leq \frac{X/n-p}{\sqrt{\frac{(X/n)(1-X/n)}{n}}} \leq z_{\frac{\alpha}{2}}\right) \approx 1-\alpha$$

We may solve as before to get the

$100(1-\alpha)\%$ confidence interval for p

$$\left(\frac{x}{n} - z_{\frac{\alpha}{2}}\sqrt{\frac{(x/n)(1-x/n)}{n}} \ , \ \frac{x}{n} + z_{\frac{\alpha}{2}}\sqrt{\frac{(x/n)(1-x/n)}{n}}\right),$$

whenever X=number of successes in n independent trials, where n is large and p is unknown.

Case Study 5.3.2 A recent poll found that 713 of 1517 respondents accepted the idea that intelligent extraterrestrials exist. What can we conclude about the proportion of all americans that believe the same thing?

ANSWER: We have $n = 1517$ and $x = 713$.

The 95% confidence interval is

$\left(\frac{x}{n} - 1.96\sqrt{\frac{(x/n)(1-x/n)}{n}} \ , \ \frac{x}{n} + 1.96\sqrt{\frac{(x/n)(1-x/n)}{n}}\right) =$

$\left(\frac{713}{1517} - 1.96\sqrt{\frac{(\frac{713}{1517})(1-\frac{713}{1517})}{1517}} \ , \ \frac{713}{1517} + 1.96\sqrt{\frac{(\frac{713}{1517})(1-\frac{713}{1517})}{1517}}\right) = (0.44, 0.50)$

CONCLUSION: IF the *true proportion* of americans who believe in extraterrestrial life is less than 0.44 or more than 0.50, it is highly unlikely that a sample proportion based on 1517 responses would be 0.47.

Example 5.3.2: Testing Random Number Generators
Suppose $Y_1, Y_2, \ldots, Y_n$ denote measurements from a continuous pdf $f_Y(y)$. Let X= number of Y_i's that are less than the median of $f_Y(y)$. If the samlple is truly random, we would expect a 95% confidence interval based on x/n to contain the value 0.5. We call this the median test.

A set of 60 computer generated samples shown in table 5.3.2 represent the exponential pdf e^{-y}, $y \geq 0$. Does this sample pass the median test?

ANSWER: First compute the median:

$$\int_0^m e^{-y}\,dy = -e^{-y}\Big|_0^m = 0.5 \Rightarrow m = 0.69315$$

Of the 60 entries in table 5.3.2, a total of 26 fall to the left of the median, so $x = 26$ and $x/n = 26/60 = 0.433$.

Let p=probability that a random observation produced by the random number generator will lie to the left of the pdf's median.

Based on these 60 observations, the 95% confidence interval for p is

$$\left(\frac{26}{60} + 1.96\sqrt{\frac{(\frac{26}{60})(1-\frac{26}{60})}{60}}, \frac{26}{60} + 1.96\sqrt{\frac{(\frac{26}{60})(1-\frac{26}{60})}{60}}\right)$$

$$= (0.308, 0.558)$$

$p = 50$ is contained in the interval, so the sample passes the test.

Binomial Dist. and Margin of Error

MARGIN OF ERROR = maximum radius of a 95% interval (usually as a percentage).

Let w:=width of a 95% confidence int. for p.
From Theorem 5.3.1,

$$\begin{aligned} w &= \frac{x}{n} + 1.96\sqrt{\frac{(x/n)(1-x/n)}{n}} \\ &\quad - \left(\frac{x}{n} + 1.96\sqrt{\frac{(x/n)(1-x/n)}{n}}\right) \\ &= 3.92\sqrt{\frac{(x/n)(1-x/n)}{n}} \end{aligned}$$

Note the largest value possible for $\frac{x}{n}(1-\frac{x}{n})$ is $\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$. Then,

$$2\,d = \max\ w = 3.92\sqrt{\frac{1}{4n}}$$

Definition 5.3.1 The **margin of error** associated with an estimate $\frac{x}{n}$, where x is the number of successes in n independent trials, is $100d$ % , where $d = \frac{1.96}{2\sqrt{n}}$.

Example 5.3.3 Modified
"40% of the Providence-area residents remembered seeing Toyota's ads on television.
The fine print of this report further states that there is a margin of error of 6% Approximately how many people were interviewed?

ANSWER:

$$0.06 = \frac{1.96}{2\sqrt{n}}$$

Then

$$\sqrt{n} = \frac{1.96}{2 \cdot 0.06} = 16.33$$

$$\Rightarrow \quad n = 16.33^2 \approx 267$$

Translation: If 100 surveys were completed with Providence residents, the true percentage answer (i.e., if every Providence resident completed the survey) would fall between 34% and 46% in 95 of the 100 surveys.

Some people mistakenly say,

"The sample is 95% accurate."

But remember, the margin of error is a measure of precision, not accuracy. Ultimately, the bottom line is that the 6% margin of error is a pretty wide margin. The real answer could just as likely be 34%, or 46%, or any other number in between.

The best estimate of the sample is that 40% of the population saw the commercials.

Choosing Sample Sizes note that smaller margin of error is achieved with larger values of n

$$\text{margin of error} = d = \frac{1.96}{2\sqrt{n}}$$

n	100	1000	10000	100000
d	0.098	0.0309903	0.0098	0.00309903

A Problem:

given the margin of error d,
compute the smallest n
needed to achieve d

Theorem 5.3.2

Let X/n be the estimator for p in binomial dist.

For X/n to have at least $100(1-\alpha)\%$ prob.

of being within a distance d of p,

the sample size should be no smaller than

$$n = \frac{z^2_{\alpha/2}}{4\,d^2}$$

Proof: Given in class.

Example 5.3.4 The proportion of children in the state, ages 0 to 14, who are lacking polio ummunization is unknown.

We wish to know how big a sample n to take to have at least a 98% probability of being within 0.05 of the true proportion p.

ANSWER: In this case,

$$100(1-\alpha) = 98$$

hence

$$\alpha = 0.02 \quad \text{and} \quad z_{\alpha/2} = 2.33$$

By theorem 2.3.2, $n = \frac{z^2_{\alpha/2}}{4\,d^2} = \frac{(2.33)^2}{4(0.05)^2} = 543$

COMMENT:

The number given in Theorem 5.3.2 is a conservative estimate. It can be lowered if additional information is available. In this case we may use the formula

$$n = \frac{z^2_{\alpha/2}\, p\,(1-p)}{d^2}$$

For example, suppose that in the previous example it is known before taking the sample that at least 80% of the children have been properly immunized. Then no more than 20% have not been properly immunized. Then

$$n = \frac{z^2_{\alpha/2}\, p\,(1-p)}{d^2} = \frac{(2.33)^2}{(0.05)^2}(0.20)(0.80) = 348$$

This is a significant reduction from the previous result (543).

Comment: Binomial or Hypergeometric? Strictly speaking, samples from surveys are drawing without replacement, is a hypergeometric process, not binomial!

However, it is ok to use the geometric model. REASONS:

- (I) $E[X/n]$ is same for both hypergeometric and binomial models.
- (II) If X is binomial, then $Var(X/n) = \frac{p(1-p)}{n}$, and if X is hypergeometric and N=total population,

$$Var(X/n) = \frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}$$

Note that

$$\frac{N-n}{N-1} \approx 1 \quad \text{if } N \text{ is much larger than } n$$