School of Mathematics and Statistics

Probability Crash Course: Large Sample Theory

Paul Hewson

Overview: This webfile is designed as a revision aid to some introductory ideas on large sample theory. It is intended to supplement a formal encouter with a text book or a set of lectures. These notes are meant to be slightly interactive, mysterious green dots, squares and boxes appear which you can click on to answer questions and check solutions.

Copyright © 2008
Last Revision Date: 23rd September 2008
paul.hewson@plymouth.ac.uk
Version 0.1

1. Introduction

We've worked with the idea that we expect an unbiased coin to come up heads around 1 time in 2. But we need to be a bit more careful about what we mean. Clearly on one throw it can only come up heads *or* tails. What we really mean is that given a large enough set of experiments, our sample mean will be *close* to the population mean. But lets tidy up our definitions.

Definition 1 *A sequence of real numbers* $\mathbb{R}_i, i = 1, 2, \ldots$ *is said to* converge *to a real number* $\mathbb{R}$ *if for any* $\epsilon > 0$ *there exists an integer* n *such that for* $i > N$ *we have:* $|\mathbb{R}_i - \mathbb{R}| < \epsilon$.

Do note the various mathematical shorthand notations we can use. We can write this as $\mathbb{R}_i \to \mathbb{R}$, and may wish to qualify that with an "as $i \to \infty$" for completeness. Or we can write it as $\lim_{i\to\infty} \mathbb{R}_i = \mathbb{R}$.

Nevertheless, while this is all very nice, it isn't a lot of use for *random* variables because sometimes it is true, sometimes not true.

We need a slightly different concept of convergence, and just to be really awkward we're going to introduce three.

Definition 2 *Convergence*

- *Convergence in probability* ($X_i \xrightarrow{P} X$)*: A sequence of random variables* X_i, $i = 1, 2, \ldots$ *converges to a random variable* X in probability *if for any* $\epsilon > 0$ *and* $\delta > 0$ *there exists an integer* N *such that for all* $i > N$ *we have:* $P(|X_n - X| < \epsilon) > 1 - \delta$

- *Convergence in mean square* ($X_i \xrightarrow{M} X$)*: A sequence* $\{X_i\}$ *converges to* X *in* mean square *if* $\lim_{i\to\infty} E[X_i - X]^2 = 0$

- *Convergence in distribution* ($X_i \xrightarrow{d} X$)*: where the distribution function* F_i *of* X_i *converges to the distribution* F *of* X *at every continuity point of* F.

Convergence in mean square implies convergence in probability which implies convergence in distribution.

2. Weak Law of Large Numbers

Now lets get to the point. If we have a sequence of random variables $\{X_i\}$, $i = 1, 2, \ldots$, we can define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (this should look rather familiar).

We want a *law of large numbers* to tell us how $\bar{X} - E[\bar{X}]$ converges to 0 in probability. This phrase used by Poisson in 1835 when discussing a 1713 paper by Bernoulli, and the weak law of large numbers is sometimes known as "Bernouilli's Theorem". But first, a small diversion.

2.1. Chebyshev's Theorem

(This happens to be an incredibly useful theorem in its own right.)

Definition 3 *Let* X *be a discrete r.v. with expected value* $\mu = E(X)$*. Let* $\epsilon > 0$ *be any positive real number. Then:*

$$P(|X - \mu| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}$$

- **Example**

Assume $E(X) = \mu$ and $Var(X) = \sigma^2$. If $\epsilon = k\sigma$, Chebyshev's inequality states:

$$P(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

i.e. for any r.v., the probability of a deviation from the mean of more than k standard deviations is $\leq \frac{1}{k^2}$. Do note that this is an important result in its own right. It might not be the most useful result in the whole world, but it is very general

Quiz

1. Let's take a quick leap and consider a continuous random variable (and hope Chebyshev's theorem also applies). In fact, let's take a standard Normal density, i.e. $X \sim N(0,1)$ If we drew a line two standard deviations either side of the mean, what proportion of the probability mass would we have (in other words what is the difference between the distribution function at $k = 2$ and $k = -2$

(a) 0.68 (b) 0.75 (c) 0.95 (d) 0.99

2. Now use Chebyshev's theorem to place a bound on the probability of a value falling 2 standard deviations either side of the mean. Compare this with the result above.

(a) 0.68 (b) 0.75 (c) 0.95 (d) 0.99

Now, before we go using this theorem to give us a proof of the weak law of large numbers, let's do some exercises with Chebyshev. It's too much fun to leave it like this.

Quiz

1. Let r denote the number of successes in n Bernoulli trials with probability of success p. Given that $Var(p) = \frac{p(1-p)}{n}$, use Chebyshev's theorem to bound the probability that $\hat{p} = \frac{r}{n}$ lies within ϵ of p

(a) $P\left(|\frac{r}{n} - p| \geq \epsilon\right) \leq \frac{p(1-p)}{\epsilon^2}$

(b) $P\left(|\frac{r}{n} - p| \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2}$

(c) $P\left(|\frac{r}{n} - p| \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon}$

(d) $P\left(|\frac{r}{n} - p| \geq \epsilon\right) \leq \frac{p(1-p)}{\epsilon}$

2. Now find the maximum possible value for $p(1-p)$:

(a) $\frac{1}{8}$ (b) $\frac{1}{4}$ (c) $\frac{1}{2}$ (d) $\frac{3}{4}$

It can be seen from Chebyshev's theorem, that as the sample size gets bigger, the difference between the sample mean and the population mean approaches zero.

2.2. The weak law of large numbers

Now we are there. We will state an important law in terms of statistical theory (we use it intuitively all the time), but it is important that it has the strength of a law, and that we can prove it.

Let $X_1, X_2, \ldots, X_n$ be an independent trials process, with finite expected value $\mu = E(X_i)$ and finite variance $\sigma^2 = Var(X_i)$. Let $\bar{X}_n = \frac{X_1, X_2, \ldots, X_n}{n}$. Then for any $\epsilon > 0$:

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$$

or even

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$$

2.3. A proof using Chebyshev's theorem

There are several ways to prove the law of large numbers. A rather nice one uses Chebyshev's theorem.

Noting that $Var(\bar{X}_n) = \frac{\sigma^2}{n}$, then $Var(n\bar{X}_n) = n\sigma^2$, and given that $E(\bar{X}_n) = \mu$ we have, for any $\epsilon > 0$:

$$P(|\bar{X}_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

so that for fixed ϵ we have:

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) \to 0$$

2.4. Strong law of large numbers

Perhaps we should also mention the strong law of large numbers:

$$P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$$

i.e. as the sample size increases, the probability that the sample mean and the population mean will be equal approches 1.

We won't give any proofs, and I'm not sure how often we need to use this particular law.

3. Central Limit Theorem

"The" (there are several, but we shall refer only to one) Central Limit theorem is one of the most useful theorems in statistical theory.

Definition 4 *The central limit theorem states: Consider S_n, the sum of n mutually independent random variables. S_n can be approximated by a normal distribution.*

This is most powerful, as the properties of the Normal distribution are very well known. For example, 95% of the probability of a *standard* normal distribution (i.e. $N(0, 1)$) lies between ± 1.96.

3.1. Example

Consider Bernoulli trials (yet more coin tossing).

If our variables are denoted by r, the sum of n Bernoulli trials we have a Binomial distribution with n and parameter p. Denoting the density here by $b(n,p,k)$, and it can be shown that

$$\lim_{n\to\infty} \sqrt{np(1-p)}b\left(n,p,(np+x\{\sqrt{np(1-p)}\}\right) = \phi(x)$$

where $\phi(x)$ is the standard Normal (i.e. with mean zero and variance 1). The only problem is that we don't know x, but we do know r. We therefore need to solve that last term for x:

If

$$r = np + x\sqrt{np(1-p)}$$

then

$$x = \frac{r-np}{\sqrt{np(1-p)}}$$

Effectively, this means that we can approximate the density of a given

Binomial variable using

$$b(n, p, k) \approx \frac{\phi(x)}{\sqrt{np(1-p)}}$$

which equals

$$\frac{1}{\sqrt{np(1-p)}}\phi\left(\frac{r-np}{\sqrt{np(1-p)}}\right)$$

Quiz

1. Find $P(S_{100} = 55)$ for $n = 100$ and $r = S_n = 55$ and $p = 0.5$

(a) 0.0480 (b) 0.0484 (c) 0.0488 (d) 0.2420

2. Now find the exact solution using the Binomial density

(a) 0.0483 (b) 0.0484 (c) 0.0485 (d) 0.0486

3.2. Example

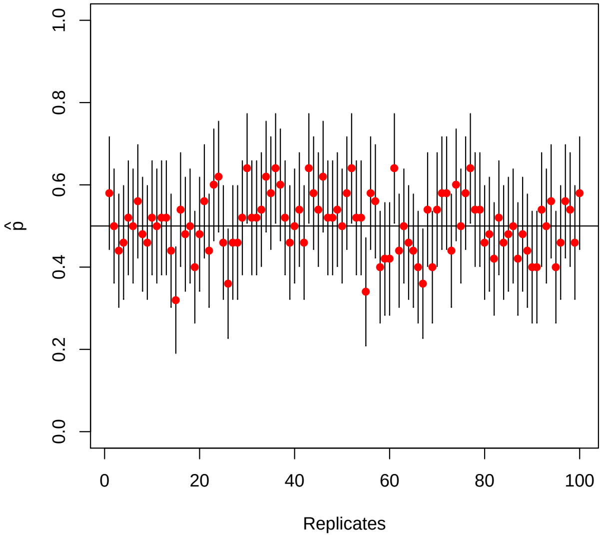
Quiz You have a sample of size $n = 9$. Your sample estimates are $\bar{x} = 12$ and $s^2 = 36$. Find the 95% confidence interval for the mean.

(a) 0 to 24 (b) 8 to 16 (c) 10 to 14 (d) 6 to 18

4. Coverage probability of a confidence interval

This is just to say what a CI really means. Confidence intervals have a quirky interpretation. If you can imagine a world in which you collect the data all over again, lots of times (perhaps you repeat the survey, perhaps you do the experiment again), you could produce your sample estimates again. A confidence interval tells you that your have $\alpha(100)\%$ confidence under these circumstances that a particular confidence interval contains the true parameter of interest.

The figure below tries to illustrate this. It's another coin tossing experiment, where we toss a coin 50 times. But we've created a world in our computer where we can conduct this experiment 100 times. We happen to know that the coin is unbiased, i.e. that $p = 0.5$. This "true" parameter is denoted by a horizontal line. The red dots illustrate the point estimates, $\hat{p}$ for each of 100 replicates. But of more interest are the calculations. These have all used the normal approximation, i.e. ± 1.96 standard errors, to produce a 95% confidence interval which is denoted by the 100 vertical lines. We therefore expect that around 95% of these confidence intervals do indeed overlap the true value of p. If you look at


1.0
0.8
0.6
0.4
0.2
0.0
$\hat{p}$
0
20
40
60
80
100
Replicates

the graph, this indeed seems to be the case. The only problem we have now, is that if you were in the real world, how would you know which of these 100 confidence intervals you were using?

Solutions to Quizzes

Solution to Quiz: If this isn't in your immediate recall, try looking at some Normal tables, or enter `>pnorm(-2) - pnorm(2)` in **R** ■

Solution to Quiz: We have from Chebyshev that $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$ so with $k = 2$ we bound the probability of exceeding 2σ as $\frac{1}{4}$, and reversing the inequality to find the bound for staying within 2 standard deviations we find that the value is 0.75.

It should be noted that Chebyshev is much more pessimistic. It says the probability mass associated with the central two standard deviations is at least 0.75, whereas using the Normal density directly suggests we could have stated this figure as 0.95. Stating something is equal to 0.95 might be more useful that stating it is greater than 0.75! However, the beauty of Chebyshev's theorem is it's generality. What happens if our data don't follow the Normal distribution? ■

Solution to Quiz: $P(S_{100} = 55 = \frac{1}{\sqrt{100 \times 0.5 \times 0.5}} \phi \left(\frac{55-50}{\sqrt{100 \times 0.5 \times (1-0.5)}} \right) = \frac{1}{5} \times 0.2420 = 0.0484$ ∎

Solution to Quiz: The most important thing to remember here is that $\bar{x} \sim N(\mu, Var(\bar{x}))$ and that we estimate $Var(\bar{x})$ by the *standard error of the mean*, that is the sample standard deviation divided by the square root of the sample size. So,

- the sample standard deviation is $\sqrt{36} = 6$ and
- the standard error of the mean is $\frac{6}{\sqrt{9}} = 2$ and so the
- 95% confidence intervals are $\bar{x} \pm 1.96 \times 2$ which is approximately 8 to 16

■