MAT 105 - Summary of Lecture 21
Tuesday, March 29, 2016

We live in the most probable of all possible worlds. (Voltaire)

Confidence intervals - examples

Recall that we normalize as follows:

(i) $X = N(\mu, \sigma^2)$ can be normalized by subtracting μ and dividing by σ (from rules about normal distributions).

$$z = \frac{X - \mu}{\sigma}$$

(ii) $X = Binomial(n, p)$ can be normalized by subtracting np and dividing by $\sqrt{np(1-p)}$ (by normal approximation to the binomial)

$$z = \frac{X - np}{\sqrt{np(1-p)}}$$

(iii) Sums of random variables can be normalized. If $S_n = X_1 + X_2 + \cdots + X_n$ with $E[X_i] = \mu$ and $Var[X_i] = \sigma^2$, then S_n is normalized by subtracting $n\mu$ and dividing by $\sqrt{n}\sigma$ (by CLT)

$$z = \frac{S_n - n\mu}{\sqrt{n}\sigma}$$

(iv) Proportions or averages of random variables can be normalized. Again, if $S_n = X_1 + X_2 + \cdots + X_n$ with $E[X_i] = \mu$ and $Var[X_i] = \sigma^2$, then the average $\overline{X} = \frac{S_n}{n}$ is normalized by subtracting μ and dividing by $\frac{\sigma}{\sqrt{n}}$ (by CLT)

$$z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

Also recall that the most commons confidence intervals are:

- the 95% confidence interval for a $N(\mu, \sigma^2)$ is $[\mu - 2\sigma\ ,\ \mu + 2\sigma]$
- the 99% confidence interval for a $N(\mu, \sigma^2)$ is $[\mu - 3\sigma\ ,\ \mu + 3\sigma]$

Homework: Think about the following problem, you do not have to hand in a solution:

To test for the "pip effect" in dice, Weldon ran the following experiment (in 1894). He rolled 12D6 exactly 26,306 times and counted the number of times 5 or 6 turned up. He recorded 106,602 times when the die landed on 5 or 6. Let

$$p = \frac{106,602}{12 \times 26306}$$

be the "estimated" probability of landing on 5 or 6.

(a) Find the 95% interval for the proportion of 5 or 6's.

(b) Use the answer in (a) to decide if the dice were fair. That is, is 1/3 in the confidence interval?

Examples – solution to Group Worksheet.

I. Suppose $X_1, X_2, \ldots$ are random variables with mean $\mu = 3$ and variance $\sigma^2 = 4$. Let $S_{100} = X_1 + \cdots + X_{100}$ be the sum of the first 100 such random variables.

(a) Approximate $\dfrac{S_{100}}{100}$

(b) Find the probability $P(260 \leq S_{100} \leq 340)$

Solution:

(a) By the Law of Large Numbers, $\dfrac{S_{100}}{100} = \mu = 3$

(b) We normalize by subtracting the mean $E[S_{100}] = 100\mu = 300$ and dividing by the standard deviation $StDev(S_{100}) = \sqrt{Var(S_{100})} = \sqrt{100\sigma^2} = \sqrt{400} = 20$:

$$P(260 \leq S_{100} \leq 340) = P\left(\frac{260-300}{20} \leq z \leq \frac{340-300}{20}\right) = P(-2 \leq z \leq 2) = \phi(2) - \phi(-2) = .9544.$$

II. We want to compute the **proportion** of heads in 100 fair coin flips.

(a) Find the probability that the proportion is less than 60%.

(b) Find the probability that the proportion is less than 40%.

Solution: We compute $\overline{X}_{100} = \dfrac{X_1 + \cdots + X_{100}}{100}$, where $X_1, X_2, \ldots$ are $Bernoulli(0.5)$, that is, they are 1 if the coin lands on heads and zero otherwise. Their mean is $\mu = 0.5$ and variance $\sigma^2 = (0.5)(1-0.5) = 0.25$, so $\sigma = 0.5$.

(a) $P(\overline{X}_{100} < 0.6) = P\left(\dfrac{\overline{X}_{100} - 0.5}{\frac{0.5}{\sqrt{100}}} < \dfrac{0.6-0.5}{\frac{0.5}{\sqrt{100}}}\right) = \phi(2) = 0.9772.$

(b) $P(\overline{X}_{100} < 0.4) = P\left(\dfrac{\overline{X}_{100} - 0.5}{\frac{0.5}{\sqrt{100}}} < \dfrac{0.4-0.5}{\frac{0.5}{\sqrt{100}}}\right) = \phi(-2) = 1 - 0.9772 = .0228.$

Another way to look at this problem is to consider $S_{100} = X_1 + \cdots + X_{100}$ and find

(a) $P(S_{100} < 60) = P\left(\dfrac{S_{100} - 100(.5)}{\sqrt{100(.25)}} < \dfrac{60 - 100(.5)}{\sqrt{100(.25)}}\right) = P(z < 2) = .9772$

(b) $P(S_{100} < 40) = P\left(\dfrac{S_{100} - 100(.5)}{\sqrt{100(.25)}} < \dfrac{40 - 100(.5)}{\sqrt{100(.25)}}\right) = P(z < -2) = 1 - .9772 = .0228.$

III. You want to check if a six-sided die is fair. You decide to count the number of times 6 comes up. Let S_n count the number of 6's in n rolls and let p be the probability that 6 comes up in a roll. If the die is indeed fair and $p = 1/6$,

(a) approximate $\dfrac{S_n}{n}$ for n large.

(b) find the 95% confidence interval for S_{100}

(c) find the 95% confidence interval for $\dfrac{S_{100}}{100}$

(d) find the 99% confidence interval for $\dfrac{S_{100}}{100}$

(e) what is the margin of error for your measurement in (d)?

Solution: Let $X_i = 1$ if the ith die lands on 6, and 0 otherwise. Then $X_i = Bernoulli(1/6)$, so $\mu = E[X_i] = 1/6$ and $\sigma^2 = Var(X_i) = (1/6)(5/6) = 5/36$.

(a) By the Law of Large Numbers, $\dfrac{S_n}{n} \approx \mu = \dfrac{1}{6}$.

(b) $S_{100} = X_1 + X_2 + \cdots + X_{100}$, so its mean, variance and standard deviation are

$$E[S_{100}] = 100E[X_1] = 100/6 = 16.66$$

$$Var(S_{100}) = 100Var(X_1) = 500/36 = 13.88,$$

$$StDev(S_{100}) = \sqrt{13.88} = 3.72.$$

Therefore, the 95% confidence interval is within 2 standard deviations of the mean:

$$[16.66 - 2(3.72), 16.66 + 2(3.72)] = [9.21, 24.12]$$

(c) $\dfrac{S_{100}}{100} = \dfrac{X_1 + X_2 + \cdots + X_{100}}{100}$, so its mean, variance and standard deviation are

$$E\left[\frac{S_{100}}{100}\right] = \frac{100E[X_1]}{100} = E[X_1] = 1/6$$

$$Var\left(\frac{S_{100}}{100}\right) = \frac{100Var(X_1)}{100^2} = \frac{5/36}{100} = 5/3600$$

$$StDev\left(\frac{S_{100}}{100}\right) = \sqrt{5/3600} = 0.0372.$$

Therefore, the 95% confidence interval is 2 standard deviations from the mean:

$$[.1666 - 2(.0372)\,,\ .1666 + 2(.0372)] = [.0921\,,\ .2412]$$

(d) Reasoning as in (c), the 99% confidence interval is 3 standard deviations from the mean:

$$[.1666 - 3(.0372)\,,\ .1666 + 3(.0372)] = [.0548\,,\ .2784]$$

(e) The margin of error in (c) is $\dfrac{2\sigma}{\sqrt{n}} = \dfrac{2(.372)}{\sqrt{100}} = .0744$. The margin of error in (d) is $\dfrac{3\sigma}{\sqrt{n}} = \dfrac{3(.372)}{\sqrt{100}} = .1116$.