Confidence Intervals
Error-Control Coding and the Binomial Distribution

Hans Georg Schaathun

Session 29 January 2015
Updated 20th January 2015

1 The Hamming Code

It is known that the $[7, 4, 3]$ Hamming code decodes correctly if and only if there is zero or one bit error in the received (7-bit) word. Two or more bit errors gives decoding error.

Exercise 1 *Calculate the error probability analytically for the* $[7, 4, 3]$ *Hamming code used on* $BSC(0.1)$*. Remember that the number of bit errors in the received 7-bit word (input to the decoder) is binomially distributed.*

Exercise 2 *Compare this error probability from Exercise 1 to your estimates in the previous session. Taking the standard error into account, are your estimates reasonable?*

2 The Standard Error

Exercise 3 *Calculate the standard error of the estimator* $\hat{p}_d$ *exactly. You can use the exact (theoretical) value for the decoding error probability* p_d *from Exercise 1 and the formula from the previous session.*

Exercise 4 *Review your sample of 100 observations of the error rate for the Hamming code over* $BSC(0.1)$ *in the previous session. How many times do you get*

1. $\hat{p} < p - \text{S.E.}(\hat{p})$?
2. $p - \text{S.E.}(\hat{p}) < \hat{p} < p + \text{S.E.}(\hat{p})$?
3. $\hat{p} > p + \text{S.E.}(\hat{p})$?

3 The theoretical distribution

For large n the binomial distribution is approximately equal to the normal distribution, and hence $\hat{p} = X/n$ has normal distribution. This fact follows from the *Central Limit Theorem*. Unless p is very close to 0 or 1, $n > 25$ qualifies as large. For a normal distribution, we have

$$P(\hat{p} < p - \text{S.E.}(\hat{p})) = 0.1587 \quad (1)$$
$$P(p - \text{S.E.}(\hat{p}) < \hat{p} < p + \text{S.E.}(\hat{p})) = 0.683 \quad (2)$$
$$P(\hat{p} > p + \text{S.E.}(\hat{p})) = 0.1587 \quad (3)$$

You can find these probabilities in a table for the standard normal distribution (or using `cdf` in Matlab).

Exercise 5 *Review Exercise 4 where you counted three events:*

1. $\hat{p} < p - \text{S.E.}(\hat{p})$?
2. $p - \text{S.E.}(\hat{p}) < \hat{p} < p + \text{S.E.}(\hat{p})$?
3. $\hat{p} > p + \text{S.E.}(\hat{p})$?

Use Equations (1)–(3) *and find the* expected number *of occurrences of each of these events.*

4 Confidence Interval

If we take Equations (1)–(3), use the estimated standard error in lieu of the exact value, and rearrange a little bit, we get

$$P(p > \hat{p} + \widehat{\text{S.E.}}(\hat{p})) = 0.1587 \quad (4)$$
$$P(\hat{p} - \widehat{\text{S.E.}}(\hat{p}) < p < \hat{p} + \widehat{\text{S.E.}}(\hat{p})) = 0.683 \quad (5)$$
$$P(\hat{p} > p + \text{S.E.}(\hat{p})) = 0.1587 P(p < \hat{p} - \widehat{\text{S.E.}}(\hat{p})) \qquad = 0.1587 \quad (6)$$
$$(7)$$

The interval $(\hat{p} - \widehat{\text{S.E.}}(\hat{p}), \hat{p} + \widehat{\text{S.E.}}(\hat{p}))$ is called a 68.3% *confidence interval* for the decoding error probability p_d. The number 68.3% is called the *confidence level.*

Exercise 6 *Using your simulation results with* $m = 100$ *tests, calculate a 68.3% confidence interval for the decoding error probability* p_d *when the* $[7, 4, 3]$ *Hamming code is used on BSC*(p)*.*

Exercise 7 *Redo Exercise 8 with* $m = 20$ *and* $m = 500$*. Compare the three confidence intervals. What do you see?*

The confidence level of 68.3% is very low, and we usually want more confidence. A $\beta = 1 - 2\alpha$ confidence interval is given by

$$(\hat{p} - z_\alpha \widehat{\mathsf{S.E.}}(\hat{p}), \hat{p} + z_\alpha \widehat{\mathsf{S.E.}}(\hat{p}))$$

The constant z_α is found in a table of the standard normal distribution. The following values are useful to remember:

1. $z_{0.1586} = 1$ for the 68.3% confidence interval
2. $z_{0.025} = 1.96$ for the 95% confidence interval
3. $z_{0.023} = 2$ for the 95.4% confidence interval
4. $z_{0.001} = 3$ for the 99.8% confidence interval

Exercise 8 *Using your simulation results with $m = 100$ tests, calculate a 95% confidence interval for the decoding error probability p_d when the $[7, 4, 3]$ Hamming code is used on $BSC(p)$.*

5 One pitfall to avoid

Consider the following to statements:

1. When you are going to calculate a 95% confidence interval for p, the probability is 95% that you get an interval which encloses p.
2. When you have calculated a 95% confidence interval (l, u) for p, the probability is 95% that $l \geq p \geq u$.

Exercise 9 *Compare the two statements above. Are they equivalent or not? Is the first statement true? Is the second statement true?*