The Binomial Distribution

Error-Control Coding and the Binomial Distribution

Hans Georg Schaathun

Session 22 January 2015

Updated 20th January 2015

In the first session, we looked at the probability distribution bit errors when a four-bit word was transmitted over BSC(p). This is an example of a binomial distribution.

The binomial distribution occurs when we conduct a number trials, where each trial is independent and has the two possible outcomes, either 'Yes' with probability p or 'No' with probability $1 - p$, and we are interested in the number X of 'Yes' events over n trials. Each trial is called a *Bernoulli trials*, and X is said to *binomially distribution* with parameters (p, n). We write $X \sim B(n, p)$.

In other words, in the last session we studied $Z \sim B(4, 0.1)$, theoretically and empirically. Now, let's turn to the general case.

1 The binomial distribution

Before you start on the following exercises, please recover your answers from the previous session.

Exercise 1 *Suppose that you send a 4-bit word on the BSC(p), for some p. Let Z be the number of bit errors. Give answers to the following questions as functions of p.*

1. *What is the probability of getting exactly z errors for $z = 0, 1, 2, 3, 4$?*
2. *What is the expected value $E(Z)$?*
3. *What is the variance $\mathsf{var}(Z)$?*

Exercise 2 *Suppose that you send a single bit word on the BSC(p), for some p. Let $Z = 1$ indicate a bit error, and $Z = 0$ indicate correct transmission. Now $Z \sim B(1, p)$.*

1. *Make a table showing the (discrete) probability distribution of Z.*
2. *What is the expected value $E(Z)$?*

Let $X_1, X_2, \ldots, X_n$ be independent stochastic variables, the we have $$\mathsf{var}\left(\sum X_i\right) = \sum \mathsf{var}(X_i), \quad (1)$$ $$E\left(\sum X_i\right) = \sum E(X_i). \quad (2)$$

Table 1: Two important formulæ.

3. *What is the variance* $\mathsf{var}(Z)$*?*

Exercise 3 *Suppose that you send an n-bit word on the BSC(p), for some p. Let Z be the number of bit errors. Give answers to the following questions as functions of p.*

1. *What is the probability of getting zero errors?*
2. *What is the probability of getting exactly one error?*
3. *What is the expected number $E(Z)$?*
4. *What is the variance* $\mathsf{var}(Z)$*?*

Exercise 4 *Compare your answers to Exercises 1 and 3. Does the special case in Exercise 1 fit with the general case in 3?*

2 The probability distribution function (PDF)

In Matlab, you can write `pdf('binom'`$, x, n, p)$ to find the probability $P(Z = x)$ where $Z \sim B(n, p)$. PDF stands for the *probability distribution function.*

Exercise 5 *Consider the transimission of a four-bit word over BSC(0.1). Use Matlab to find the probabilities of zero, one, two, three, or four bit errors. Compare the result to your own manual calculations from previous exercises.*

Now, let's see if we can visualise the distribution for large n. Let's keep $p = 0.1$ for now. We want to plot the probability $P(Z = z)$ for z ranging from 0 to n. To do that, we need to generate a vector (1-D matrix) of z values, and the corresponding vector of probabilities. In matlab $Z = [0 : n]$ will make Z a vector of all values $0, 1, \ldots, n$. If n is very large, you can take every tenth value by writing $Z = [0 : 10 : n]$. Or choose a different value in lieu of 10. Once you have your `Z` vector, you can calculate all the probabilities in one call, as `P = pdf('binom',Z,n,0.1)`. You can create the plot with `plot(Z,P)`.

Exercise 6 *Plot the probability distribution function of $Z \sim B(100, 0.1)$.*

Note that Z ranges from 0 to n. To be able to compare distributions for diffent n, it is useful to normalise and consider the stochastic variable $X = Z/n$ instead. That really means that you replace your `Z` vector by `X = Z/n`. Otherwise the approach is identical to the above.

Exercise 7 *Plot the probability distribution function of* $X = Z/100$ *where* $Z \sim B(100, 0.1)$.

Finally, let's compare different distributions. You can plot multiple functions in the same figure by issuing the `hold` command after the first call to `plot`. You can also plot in different colours using additional arguments to `plot`. See the help page.

Exercise 8 *Plot and compare the probability distribution functions of the stochastic variable* X/n *where* $X \sim B(n, 0.05)$ *for* $n = 10, 100, 1000$.

What observation can you make?

Judging from the plot, what appears to be the mean value of X*?*

3 The cummulative distribution function (CDF)

The `pdf` function (for a discrete distribution) gives you the probability $P(X = x)$. Another important function is `cdf` (*Cummulative Distribution Function*) which gives the probability $P(X \leq x)$.

Exercise 9 *Suppose you send a word of 1000 bits over a BSC bit bit error probability* $p = 0.02$. *What is the probability of getting at most ...*

1. 2% *bit errors?*
2. 5% *bit errors?*

(Use matlab to find the answer.)