Introduction to Statistics
Binomial distribution

Grzegorz Chrupała

Saarland University

Binomial distribution

$$\mathrm{b}(r; n, p) = \binom{n}{r} p^r (1-p)^{n-r}$$

where

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}, 0 \leq r \leq n$$

- $b(r, n, p)$ describes the probability of getting exactly r successes in n trials if the probability of success in an individual trial is p
- $\binom{n}{r}$ is the number of different orders in which we can get r successes in n trials
- Each attempt is independent, so we multiply p r times (successes) and $(1-p)$, $n-r$ times (failures)
- What is the probability of getting **at most** r successes?

Binomial distribution

$$\mathrm{b}(r; n, p) = \binom{n}{r} p^r (1-p)^{n-r}$$

where

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}, \; 0 \leq r \leq n$$

- $b(r, n, p)$ describes the probability of getting exactly r successes in n trials if the probability of success in an individual trial is p
- $\binom{n}{r}$ is the number of different orders in which we can get r successes in n trials
- Each attempt is independent, so we multiply p r times (successes) and $(1-p)$, $n-r$ times (failures)
- What is the probability of getting **at most** r successes?

$$\sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$$

Binomial test example

- We have made an improvement to our POS tagging model.
- We run the old model and the new model on test sentences.
- The accuracy of the new model is better, but
  - Is it because the system is better? If we repeated the experiment on many other test sentences, would be also get improved accuracy?
  - Or maybe we got an improvement by chance

Null hypothesis

- Use binomial distribution to answer this question
- Focus on the tokens (words) where one of the models makes a mistake and the other gets the right answer
- There are 10 such cases. In 7 cases the new system is better.
- Assume that the new system is actually no better, and that the chance of it being better on any one word is pure chance, 0.5. This is the null hypothesis.
  - How likely are we to get **at least** 7 out of 10 better, given the null hypothesis?
  - How about 55 out of 100? 550 out of 1000?

Null hypothesis

- Use binomial distribution to answer this question
- Focus on the tokens (words) where one of the models makes a mistake and the other gets the right answer
- There are 10 such cases. In 7 cases the new system is better.
- Assume that the new system is actually no better, and that the chance of it being better on any one word is pure chance, 0.5. This is the null hypothesis.
  - How likely are we to get **at least** 7 out of 10 better, given the null hypothesis?
  - How about 55 out of 100? 550 out of 1000?
- Two-tailed test:
  - Actually we should consider both getting **at least 7 out of 10** or **at most 3 out of 10**