Some uses of the binomial distribution

Daniel Hsu

October 7, 2023

1 Test error rate

Suppose P is the probability distribution over $\mathcal{X} \times \mathcal{Y}$ of interest. The error rate of a classifier $f \colon \mathcal{X} \times \mathcal{Y}$ is defined by

$$\mathrm{err}[f] = \Pr(f(X) \neq Y).$$

where $(X, Y) \sim P$.

Suppose you have a classifier $f \colon \mathcal{X} \to \mathcal{Y}$ and test data

$$(\tilde{X}^{(1)}, \tilde{Y}^{(1)}), \dots, (\tilde{X}^{(m)}, \tilde{Y}^{(m)}) \sim_{\text{i.i.d.}} P,$$

and you would like to estimate the error rate of f. Let S denote the number of test examples on which f makes a prediction error, i.e.,

$$S = \sum_{i=1}^{m} \mathbb{1}\{f(\tilde{X}^{(i)}) \neq \tilde{Y}^{(i)}\}.$$

Then the test error rate of f, which we'll denote by $\widetilde{\mathrm{err}}[f]$, is equal to

$$\widetilde{\mathrm{err}}[f] = \frac{S}{m}.$$

The distribution of S is $\mathrm{Binomial}(m, \theta)$, where $\theta = \mathrm{err}[f]$. Therefore

$$\mathbb{E}(S) = m\theta, \quad \mathrm{var}(S) = m\theta(1 - \theta),$$

and

$$\mathbb{E}(\widetilde{\mathrm{err}}[f]) = \theta = \mathrm{err}[f], \quad \mathrm{stddev}(\widetilde{\mathrm{err}}[f]) = \sqrt{\frac{\theta(1-\theta)}{m}} = \sqrt{\frac{\mathrm{err}[f](1 - \mathrm{err}[f])}{m}}.$$

As $m \to \infty$, the central limit theorem implies that the binomial distribution converges to a normal distribution in a certain sense. In particular:

$$\sqrt{m} \cdot \frac{\widetilde{\mathrm{err}}[f] - \mathrm{err}[f]}{\sqrt{\mathrm{err}[f](1 - \mathrm{err}[f])}} \longrightarrow \mathrm{N}(0, 1).$$

Since the normal distribution contains about 95% of its probability mass within two standard deviations of its mean, we have (for large m), with probability $\approx 95\%$,

$$\mathrm{err}[f] \leq \widetilde{\mathrm{err}}[f] + 2\sqrt{\frac{\mathrm{err}[f](1 - \mathrm{err}[f])}{m}},$$
$$\mathrm{err}[f] \geq \widetilde{\mathrm{err}}[f] - 2\sqrt{\frac{\mathrm{err}[f](1 - \mathrm{err}[f])}{m}}.$$

When these two inequalities hold, we can deduce upper- and lower-bounds on $\mathrm{err}[f]$ in terms of $\widetilde{\mathrm{err}}[f]$ and m by solving a quadratic equation. See this GitHub gist for how this can be done.

2 Is heads or tails is more likely?

Suppose you have a coin that you suspect is biased, and you would like to determine whether heads or tails is more likely. Letting θ denote the probability of heads:

- heads is more likely than tails if $\theta > 1/2$;
- tails is more likely than heads if $\theta < 1/2$.

If $\theta = 1/2$, we are fine with picking either heads or tails.

Without knowledge of θ, we attempt to make the determination based on the results of tossing the coin several times. Let S denote the number of tosses that are heads in n independent tosses of the coin. Our guess is

- heads if $S > n/2$;
- tails if $S \leq n/2$.

Suppose $\theta > 1/2$, so heads is more likely than tails. Our guess is incorrect if $S \leq n/2$. What is the probability of this event? In particular, how does it depend on the number of tosses?

For simplicity let us assume that n is even. We know that $S \sim \text{Binomial}(n, \theta)$, so using the probability mass function for S, we have

$$
\begin{aligned}
\Pr(S \leq n/2) &= \sum_{k=0}^{n/2} \binom{n}{k} \theta^k (1-\theta)^{n-k} \\
&= \sum_{k=0}^{n/2} \binom{n}{k} 2^{-n} \left(\frac{\theta}{1/2}\right)^k \left(\frac{1-\theta}{1/2}\right)^n \left(\frac{1/2}{1-\theta}\right)^k \\
&\leq \sum_{k=0}^{n/2} \binom{n}{k} 2^{-n} \left(\frac{\theta}{1/2}\right)^{n/2} \left(\frac{1-\theta}{1/2}\right)^n \left(\frac{1/2}{1-\theta}\right)^{n/2} \\
&= (4\theta(1-\theta))^{n/2} \sum_{k=0}^{n/2} \binom{n}{k} 2^{-n} \\
&\leq (4\theta(1-\theta))^{n/2} \sum_{k=0}^{n} \binom{n}{k} 2^{-n} \\
&= (4\theta(1-\theta))^{n/2}.
\end{aligned}
$$

The first inequality above uses the facts that $\theta > 1/2$, and that each term in the summation has $k \leq n/2$. The final step uses the binomial theorem. Notice that, for any $\theta \neq 1/2$,

$$4\theta(1-\theta) < 1.$$

Hence

$$\Pr(S \leq n/2) \leq (4\theta(1-\theta))^{n/2} = \exp(-cn)$$

for

$$c = \frac{1}{2} \ln\left(\frac{1}{4\theta(1-\theta)}\right) > 0.$$

The probability that our guess is incorrect is exponentially small in the number of tosses n.

The case where $\theta < 1/2$ can be handled in a completely symmetric manner.