Bernoulli Law of Large Numbers and Weierstrass' Approximation Theorem

Márton Balázs* and Bálint Tóth*

October 13, 2014

This little write-up is part of important foundations of probability that were left out of the unit Probability 1 due to lack of time and prerequisites. Here we give an elementary proof of the Bernoulli Weak Law of Large Numbers. As a corollary, we prove Weierstrass' Approximation Theorem regarding Bernstein's polynomials.

We need the notion of the *mode* of a discrete distribution: this is simply the most likely value(s) of our random variable. In other words, this is the value(s) x_i where the mass function $p_X(x_i)$ is maximal.

Proposition 1 *The mode of the Binomial distribution of parameters n, p is $\lfloor (n+1)p \rfloor$ (lower integer part). If $(n+1)p$ happens to be an integer number, then $(n+1)p-1$ and $(n+1)p$ are both modes of this distribution.*

Proof. Let $0 < i \le n$, then

$$\frac{p_X(i)}{p_X(i-1)} = \frac{\binom{n}{i} p^i (1-p)^{n-i}}{\binom{n}{i-1} p^{i-1} (1-p)^{n-i+1}} = \frac{(n-i+1)p}{i(1-p)}.$$

Therefore, the following are equivalent:

$$\begin{aligned} p_X(i) &\ge p_X(i-1) \\ \frac{p_X(i)}{p_X(i-1)} &\ge 1 \\ \frac{(n-i+1)p}{i(1-p)} &\ge 1 \\ (n+1)p &\ge i \\ \lfloor (n+1)p \rfloor &\ge i. \end{aligned}$$

This implies that $p_X(i)$ increases until i reaches the value $\lfloor (n+1)p \rfloor$, after which we see a decrease of the mass function. We also see that $p_X(i) = p_X(i-1)$ happens if and only if $i = (n+1)p$ (which requires $(n+1)p$ to be an integer), in this case we have two modes. □

Recall that the Binomial(n, p) distribution counts the number of successes in a given n number of trials. When this number is large, it is natural to expect that the proportion of successes approximates the success probability p. This is the content of the Bernoulli Law of Large Numbers:

Theorem 2 (Bernoulli Law of Large Numbers) *Fix $0 < p < 1$, and for given n let $X \sim$ Binom(n, p). Then for all $\varepsilon > 0$*

$$\lim_{n\to\infty} \mathbf{P}\Big\{ \Big| \frac{X}{n} - p \Big| > \varepsilon \Big\} = 0.$$

The Law of Large Numbers can be proved in various versions with various methods, see a general one at the end of the Probability 1 slides, or stronger versions in more advanced texts on probability.

Proof. For the sake of simplicity we use the abbreviation $q = 1 - p$. Let $r \ge (n+1)p$ and $k \ge 1$ be integers. Then a simple computation for the ratio of the Binomial mass functions shows

$$\frac{p_X(r+k)}{p_X(r+k-1)} = \frac{n-r-k+1}{r+k} \cdot \frac{p}{q} \le \frac{n-r}{r+k} \cdot \frac{p}{q} \le \frac{n-r}{r} \cdot \frac{p}{q} =: K.$$

This number K is bounded by 1, since

$$K = \Big(\frac{n}{r} - 1 \Big) \cdot \frac{p}{q} \le \Big(\frac{n}{(n+1)p} - 1 \Big) \cdot \frac{p}{q} = \frac{n-np-p}{(n+1)p} \cdot \frac{p}{q} < \frac{n-np+1-p}{(n+1)p} \cdot \frac{p}{q} = \frac{(n+1)\cdot(1-p)}{(n+1)p} \cdot \frac{p}{q} = 1.$$

*University of Bristol / Budapest University of Technology and Economics

With this, we have

$$\mathbf{P}\{X \geq r\} = \sum_{k=0}^{n-r} p_X(r+k) = \sum_{k=0}^{n-r} p_X(r) \cdot \underbrace{\frac{p_X(r+1)}{p_X(r)} \cdot \frac{p_X(r+2)}{p_X(r+1)} \cdots \frac{p_X(r+k)}{p_X(r+k-1)}}_{k \text{ terms}} \leq$$

$$\leq \sum_{k=0}^{n-r} p_X(r) \cdot K^k = p_X(r) \cdot \frac{1-K^{n-r+1}}{1-K} \leq \frac{p_X(r)}{1-K}.$$

Next is an estimate for $p_X(r)$, for which we use Proposition 1. By our assumption $r \geq (n+1)p$, therefore $r \geq \lfloor (n+1)p \rfloor$, and the mass function decreases between $\lfloor (n+1)p \rfloor$ and r. Thus for all $\lfloor (n+1)p \rfloor \leq i \leq r$ we have $p_X(r) \leq p_X(i)$, and a rather crude estimate gives

$$1 = \sum_{i=0}^{n} p_X(i) \geq \sum_{i=\lfloor (n+1)p \rfloor}^{r} p_X(i) \geq \sum_{i=\lfloor (n+1)p \rfloor}^{r} p_X(r) = \big(r - \lfloor (n+1)p \rfloor + 1\big) \cdot p_X(r) \geq$$

$$\geq \big(r - (n+1)p + 1\big) \cdot p_X(r) \geq (r - np) \cdot p_X(r), \text{ that is}$$

$$p_X(r) \leq \frac{1}{r-np}.$$

We proceed from here and the definition of K with estimating the probability:

$$\mathbf{P}\{X \geq r\} \leq \frac{1}{r-np} \cdot \frac{1}{1 - \frac{n-r}{r} \cdot \frac{p}{q}} = \frac{rq}{(r-np)^2}.$$

With ε and p given, $n\varepsilon > p$ will hold for all large enough n's, and we can make the choice $r = \lceil np + n\varepsilon \rceil$ (upper integer part) in the previous estimate:

$$\mathbf{P}\Big\{\frac{X}{n} - p > \varepsilon\Big\} = \mathbf{P}\{X > np + n\varepsilon\} \leq \mathbf{P}\{X \geq \lceil np + n\varepsilon \rceil\} \leq \frac{\lceil np + n\varepsilon \rceil q}{\big(\lceil np + n\varepsilon \rceil - np\big)^2} \leq \frac{(np + n\varepsilon + 1)q}{n^2\varepsilon^2} =$$

$$= \frac{pq}{\varepsilon^2 n} + \frac{q}{\varepsilon n} + \frac{q}{\varepsilon^2 n^2}.$$

To finish the proof we also need a lower bound on X. To do this, we notice that $Y := n - X$, the number of failures, is Binom$(n,\, q)$ distributed. All the above applies to this variable, and we can write

$$\mathbf{P}\Big\{\frac{X}{n} - p < -\varepsilon\Big\} = \mathbf{P}\Big\{\frac{n-Y}{n} - p < -\varepsilon\Big\} = \mathbf{P}\Big\{\frac{-Y}{n} + 1 - p < -\varepsilon\Big\} = \mathbf{P}\Big\{\frac{Y}{n} - q > \varepsilon\Big\} = \frac{pq}{\varepsilon^2 n} + \frac{p}{\varepsilon n} + \frac{p}{\varepsilon^2 n^2}.$$

finally,

$$\mathbf{P}\Big\{\Big|\frac{X}{n} - p\Big| > \varepsilon\Big\} \leq \mathbf{P}\Big\{\frac{X}{n} - p > \varepsilon\Big\} + \mathbf{P}\Big\{\frac{X}{n} - p < -\varepsilon\Big\} \leq \frac{2pq}{\varepsilon^2 n} + \frac{1}{\varepsilon n} + \frac{1}{\varepsilon^2 n^2} \leq \frac{1}{2\varepsilon^2 n} + \frac{1}{\varepsilon n} + \frac{1}{\varepsilon^2 n^2}, \qquad (1)$$

where in the last step we used that $pq = p(1-p)$ can never be larger than $1/4$. □

Notice that we actually proved more than the statement of the theorem: for any $\varepsilon > p/n$ (1) holds. As an example, we can choose $\varepsilon = \varepsilon(n)$ to be a function of n, we just have to make sure it decreases slow enough for $\varepsilon(n) \cdot \sqrt{n} \underset{n\to\infty}{\longrightarrow} \infty$ to hold. Then we still have

$$\mathbf{P}\Big\{\Big|\frac{X}{n} - p\Big| > \varepsilon(n)\Big\} \underset{n\to\infty}{\longrightarrow} 0.$$

As an application we prove Weierstrass' approximation theorem:

Theorem 3 (Weierstrass' approximation theorem) *Let $f : [0,\, 1] \to \mathbb{R}$ be a continuous function. Then for all $\varepsilon > 0$ there exist $n < \infty$ and a polynomial $B_n(x)$ of degree n, such that*

$$\sup_{0 \leq x \leq 1} |f(x) - B_n(x)| < \varepsilon.$$

Proof. Given $x \in [0,\, 1]$, let $X \sim$ Binom$(n,\, x)$, and define the *Bernstein-polynomial* of degree n by

$$B_n(x) := \mathbf{E}\Big[f\Big(\frac{X}{n}\Big)\Big] = \sum_{i=0}^{n} f\Big(\frac{i}{n}\Big)\binom{n}{i}(1-x)^{n-i}x^i.$$

f is continuous on a closed interval, hence it is bounded, and it is also uniformly continuous by Heine's theorem. Therefore with $\varepsilon/2$ we have $\delta > 0$ such that $|f(x) - f(y)| \le \varepsilon/2$ for all $0 \le x,\, y \le 1$, $|x - y| \le \delta$. With this δ,

$$\begin{aligned}\bigl|f(x) - B_n(x)\bigr| = \bigl|f(x) - \mathbf{E}\bigl[f(X/n)\bigr]\bigr| = \bigl|\mathbf{E}\bigl[f(x) - f(X/n)\bigr]\bigr| \le \\ \le \bigl|\mathbf{E}\bigl[(f(x) - f(X/n)) \cdot \mathbf{1}\{|x - X/n| > \delta\}\bigr]\bigr| + \bigl|\mathbf{E}\bigl[(f(x) - f(X/n)) \cdot \mathbf{1}\{|x - X/n| \le \delta\}\bigr]\bigr| \le \\ \le 2M \cdot \mathbf{P}\{|x - X/n| > \delta\} + \varepsilon/2,\end{aligned}$$

where in the last step we bounded f by its maximum M in $[0,\, 1]$, used uniform continuity in the second term, and the fact that an indicator never exceeds 1. Using (1) we obtain that

$$\bigl|f(x) - B_n(x)\bigr| \le 2M\Bigl[\frac{1}{2\delta^2 n} + \frac{1}{\delta n} + \frac{1}{\delta^2 n^2}\Bigr] + \varepsilon/2.$$

For a given ε and the appropriately chosen δ we have a large enough n that makes the right hand-side smaller than ε, and this finishes the proof. $\square$