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 $0 < i \leq n$, 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:

$$p_X(i) \geq p_X(i-1)$$

$$\frac{p_X(i)}{p_X(i-1)} \geq 1$$

$$\frac{(n-i+1)p}{i(1-p)} \geq 1$$

$$\frac{(n+1)p}{i} \geq i$$

$$\lfloor (n+1)p \rfloor \geq i.$$  

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 $X \sim \text{Binom}(n, p)$. Then for all $\varepsilon > 0$

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

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 $r \geq (n+1)p$ and $k \geq 1$ be integers. Then a simple computation for the ratio of the Binomial mass functions shows

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

This number $K$ is bounded by 1, since

$$K = \binom{n}{r} \cdot \frac{p}{q} \leq \binom{n}{(n+1)p} \cdot \frac{p}{q} = \frac{n-np}{(n+1)p} \cdot \frac{p}{q} \leq \frac{n-np+1-p}{(n+1)p} \cdot \frac{p}{q} = \frac{(n+1)(1-p)}{(n+1)p} \cdot \frac{p}{q} = 1.$$

---

*University of Bristol / Budapest University of Technology and Economics
With this, we have

$$\Pr[X \geq r] = \sum_{k=0}^{n-r} p_X(r + k) = \sum_{k=0}^{n-r} p_X(r) \cdot \frac{p_X(r+1)}{p_X(r)} \cdot \frac{p_X(r+2)}{p_X(r+1)} \cdots \frac{p_X(r+k)}{p_X(r+k-1)} \leq \sum_{k=0}^{n-r} p_X(r) \cdot K^k = p_X(r) \cdot \frac{1 - K^{n-r+1}}{1 - K} \leq p_X(r).$$

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

$$1 = \sum_{i=0}^{n} p_X(i) \geq \sum_{i=\lceil (n+1)p \rceil}^{r} p_X(i) \geq \sum_{i=\lceil (n+1)p \rceil}^{r} p_X(r) = (r - \lceil (n+1)p \rceil) \cdot p_X(r) \geq (r - (n+1)p + 1) \cdot p_X(r) \geq (r - np) \cdot p_X(r),$$

that is

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

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

$$\Pr[X \geq r] \leq \frac{1}{r - np} \cdot \frac{1}{1 - \frac{r}{n} \cdot \frac{\varepsilon}{q}} = \frac{rq}{(r - np)^2}.$$

With $\varepsilon$ 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:

$$\Pr\left\{ \frac{X}{n} - p > \varepsilon \right\} = \Pr\{X > np + n\varepsilon\} \leq \Pr\{X \geq \lceil np + n\varepsilon \rceil\} \leq \frac{\lceil np + n\varepsilon + 1 \rceil}{\lceil np + n\varepsilon \rceil - np} = \frac{np + n\varepsilon + 1}{n\varepsilon^2} = \frac{pq}{\varepsilon^2 n} + \frac{q}{\varepsilon n} + \frac{q}{\varepsilon^2 n^2}.$$

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

$$\Pr\left\{ \frac{X}{n} - p < -\varepsilon \right\} = \Pr\left\{ \frac{n - Y}{n} - p < -\varepsilon \right\} = \Pr\left\{ \frac{Y}{n} + 1 - p < \varepsilon \right\} = \Pr\left\{ \frac{Y}{n} - q > \varepsilon \right\} = \frac{pq}{\varepsilon^2 n} + \frac{p}{\varepsilon n} + \frac{p}{\varepsilon^2 n^2},$$

finally,

$$\Pr\left\{ \left| \frac{X}{n} - p \right| > \varepsilon \right\} \leq \Pr\left\{ \frac{X}{n} - p > \varepsilon \right\} + \Pr\left\{ \frac{X}{n} - p < -\varepsilon \right\} \leq \frac{2pq}{\varepsilon^2 n} + \frac{1}{\varepsilon n} + \frac{1}{\varepsilon^2 n^2} \leq \frac{1}{2\varepsilon^2 n} + \frac{1}{\varepsilon n} + \frac{1}{\varepsilon^2 n^2}, \quad (1)$$

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 $\varepsilon(n) \cdot \sqrt{n} \to \infty$ to hold. Then we still have

$$\Pr\left\{ \left| \frac{X}{n} - p \right| > \varepsilon(n) \right\} \to 0.$$

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 $X \sim \text{Binom}(n, x)$, and define the *Bernstein-polynomial* of degree $n$ by

$$B_n(x) := \mathbb{E}\left[f\left(\frac{X}{n}\right)\right] = \sum_{i=0}^{n} f\left(\frac{i}{n}\right) \binom{n}{i} (1-x)^{n-i} x^i.$$
$f$ 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 $|f(x) - f(y)| \leq \varepsilon/2$ for all $0 \leq x, y \leq 1$, $|x - y| \leq \delta$. With this $\delta$,

$$
|f(x) - B_n(x)| = |f(x) - E[f(X/n)]| = |E[f(x) - f(X/n)]| \leq \\
\leq |E[(f(x) - f(X/n)) \cdot 1\{|x - X/n| > \delta\}]| + |E[(f(x) - f(X/n)) \cdot 1\{|x - X/n| \leq \delta\}]| \leq \\
\leq 2M \cdot P\{|x - X/n| > \delta\} + \varepsilon/2,
$$

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

$$
|f(x) - B_n(x)| \leq 2M \left[ \frac{1}{2\delta^2 n} + \frac{1}{\delta n} + \frac{1}{\delta^2 n^2} \right] + \varepsilon/2.
$$

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