Elements of Discrete Probability
1. The Central Limit Theorem and the Normal Distribution

1 The law of large numbers (from probabilities to frequencies)

Let f be a random variable with known variance σ^2 but unknown expectation, and let $f_1, f_2, \ldots, f_n$ be a sequence of independent observations (samples) of this random variable. Such sequence is analogous to a sequence of repeated trials, with the difference that f is, in general, not Bernoulli; more appropriate is to think of f as a measurement of some physical quantity, with an instrument of which we know the precision. Equivalently, the $f_1, f_2, \ldots, f_n$ can be interpreted as n independent random variables with a common distribution. Let $E[f] = \mu$. We now calculate the arithmetic average

$$m = \frac{f_1 + f_2 + \ldots + f_n}{n}$$

which is itself a random variable.

Theorem 1 (Law of large numbers) *Let f be a random variable with expectation $E[f] = \mu$ and variance $var[Z] = \sigma^2$. Then:*

$$\lim_{n\to\infty} \Pr\left(\left|\frac{f_1 + \ldots + f_n}{n} - \mu\right| \geq \epsilon\right) = 0$$

Proof:

By the linearity of expectation

$$E[m] = \frac{1}{n}(E[f_1] + E[f_2] + \ldots + E[f_n]) = \frac{n\mu}{n} = \mu$$

since $f_1, f_2, \ldots, f_n$ have the same distribution and $E[f_j] = \mu$.

In addition, since the $f_1, f_2, \ldots, f_n$ are independent

$$var[m] = var\left[\frac{f_1}{n}\right] + \ldots + var\left[\frac{f_n}{n}\right] = n \cdot var\left[\frac{f}{n}\right]$$

We now observe that

$$var\left[\frac{f}{n}\right] = E\left[\left(\frac{f-\mu}{n}\right)^2\right] = \frac{E[(f-\mu)^2]}{n^2} = \frac{var[f]}{n^2}$$

and obtain

$$var[m] = n\frac{var[n]}{n^2} = \frac{\sigma^2}{n}$$

We now apply the Chebyshev's inequality to variable m and obtain

$$\Pr\left(\left|\frac{f_1 + \ldots + f_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{var[m]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

and the theorem follows when we let n grow to ∞, because σ^2/ϵ^2 is a constant. $\square$

The law of large numbers is one of the most important results of probability theory. In fact, m is the mean of the samples whereas μ is the mean (expectation) of the process that generates the samples, i.e., the mean of the population. Thus, the previous result can be rewritten as

$$\lim_{n\to\infty} \Pr(\text{mean(sample)} - \text{mean(population)} \geq \epsilon) = 0$$

which is rephrased as

$$\lim_{n\to\infty} \text{mean(sample)} = \text{mean(population)}$$

in probability, and which ties frequencies to probabilities in a precise and beautiful way and validates our intuition/motivation for probability theory.

2 The normal distribution and the central limit theorem

Consider a sequence of n Bernoulli trials with probability p (n independent tossings of a possibly unfair coin). We know that the (random variable) "number of heads" j obeys a binomial distribution

$$B(n,p;j) = \binom{n}{j} p^j q^{n-j}$$

If we plot this distribution, we'll observe an approximately bell-shaped configuration of vertical bars (or spikes), centered aouund the mean np and with an apparent spread $\approx 5\sqrt{npq}$ (where npq, we recall, is the variance of j) (see Figure 1, upper envelope).

We now double n and imagine to produce the resulting plot. Of course, the value of the mean doubles as well and the apparent spread increases by a factor $\sqrt{2}$ (since it grows with the standard deviation $\sqrt{npq}$). For a fixed value of p, we would like to compare the histograms of distributions corresponding to different values of n so their mean values coincides and their spreads appear identical. The first objective is achieved by replacing the value j with its signed distance from its mean $(j - np)$, effectively translating the histogram so that the mean is brought to coincide with 0. The second objective (identical spreads) is achieved by "normalizing" the standard deviation to to the value 1, i.e., by dividing the number $j - \mu$ by $\sqrt{npq}$. In other words, the result of these two modifications is to replace the integer j with the normalized value(a real number)

$$x = \frac{j - np}{\sqrt{npq}}, \tag{1}$$

In Figure 1 we have the normalized histograms corresponding to n and $4n$.

With this convention, if we compare the diagrams of distributions corresponding to different values of n for fixed p, we shall observe that the number of spikes between the mean and 1 standard deviation grows like the $\sqrt{n}$: for example, for $p = 0.5$ there are about 5 spikes for $n = 100$, 10 spikes for $n = 400$, and so on, while the shape of the outer envelope of the diagrams appears not to change (except for the fact that the values decrease like $1/\sqrt{n}$). If we now let the number n grow unbounded, the spike diagram will become denser and denser, and when we let n become infinite the distribution will no longer be "discrete": it will become "continuous"!

We wish to analyze this behavior. Again we refer to the normalized variable defined above. With the substitution $j = \sqrt{npq}x + np$ (see (1) above), the original binomial distribution, multiplied by the normalizing factor $\sqrt{npq}$, becomes:

$$\sqrt{pnq}B(n,p,j) = \sqrt{npq}\binom{n}{\sqrt{npq}x + np} p^{np + x\sqrt{npq}} q^{nq - x\sqrt{npq}}$$

We then have the following, very important, theorem:

Theorem 2 *(***Central limit theorem for the binomial distribution***) For fixed x and p we have:*

$$\lim_{n\to\infty} \sqrt{npq}B(n,p;j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \Phi(x)$$

The function $\Phi(x)$ is (the density function of) the **Normal distribution** *with mean 0 and variance 1*, which is perhaps the most important probability (continuous) distribution.

An elementary proof of this theorem will be given separately below. The significance of the above result is the following: Note that the number j of heads can

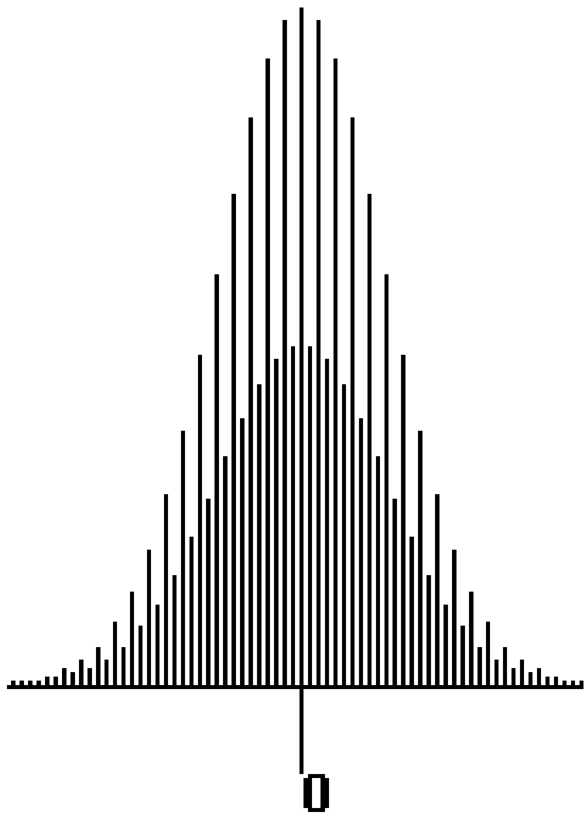

Figure 1: Superposed diagrams of the spread-normalized binomial distributions for n and $4n$.

really be viewed as the sum of n independent identically distributed random variables $x_1, x_2, \ldots, x_n$ each of which has value 1 for "head" and 0 for "tail". From the definitions, the common mean and variance of these variables are p and pq, respectively. With this interpretation, the above theorem states that as the number n grows, the sum of these variables tends to be normally distributed with mean np and variance npq.

A most remarkable result, stated here without proof, is that the central limit theorem can be generalized to arbitrary independent random variables (not just Bernoulli trials) as follows:

Theorem 3 *(**Central limit theorem**) Let $x_1, x_2, \ldots, x_n$ be independent identically distributed random variables, whose distribution has finite mean μ and variance σ^2. Then, as n grows, the distribution of $\sum_{j=1}^{n} x_j$ tends to the normal distribution with* *mean $n\mu$ and variance $n\sigma^2$.*

Derivation of the limiting binomial distribution

Unfortunately, the binomial coeffcients are awkward analytical objects, since factorials don't lend themselves to easy manipulations. Luckily, there is an excellent approximation of the factorial in terms of elementary functions, the *Stirling approximation*:

$$n! \approx \sqrt{2\pi n}n^n e^{-n}$$

(For example, $6! = 720$ and $\sqrt{2\pi 6}6^6 e^{-6} \approx 710$.) Using this approximation, the above distribution is so transformed:

$$\sqrt{npq}\binom{n}{\sqrt{npq}x + np} p^{np+x\sqrt{npq}} q^{nq-x\sqrt{npq}}$$

$$\begin{aligned}
&= \sqrt{npq}\sqrt{\frac{n}{2\pi(np+x\sqrt{npq})(nq-x\sqrt{npq})}}\left(\frac{np}{np+x\sqrt{npq}}\right)^{np+x\sqrt{npq}}\left(\frac{nq}{nq-x\sqrt{npq}}\right)^{nq-x\sqrt{npq}} \\
&= \sqrt{pnq}\frac{1}{\sqrt{2\pi(npq)(\sqrt{\frac{p}{q}}+\frac{x}{\sqrt{n}})(\sqrt{\frac{q}{p}}-\frac{x}{\sqrt{n}})}}\left(1+\frac{x}{\sqrt{n}}\sqrt{\frac{q}{p}}\right)^{-np-x\sqrt{npq}}\left(1-\frac{x}{\sqrt{n}}\sqrt{\frac{p}{q}}\right)^{-nq+x\sqrt{npq}}
\end{aligned}$$

We shall consider separately the two terms

$$\sqrt{pnq}\frac{1}{\sqrt{2\pi(npq)(\sqrt{\frac{p}{q}}+\frac{x}{\sqrt{n}})(\sqrt{\frac{q}{p}}-\frac{x}{\sqrt{n}})}} \quad \text{and} \quad \left(1+\frac{x}{\sqrt{n}}\sqrt{\frac{q}{p}}\right)^{-np-x\sqrt{npq}}\left(1-\frac{x}{\sqrt{n}}\sqrt{\frac{p}{q}}\right)^{-nq+x\sqrt{npq}}$$

and begin with the first one. We wish to compute the limit as n goes to infinity:

$$\begin{aligned}
\lim_{n\to\infty}\sqrt{pnq}\frac{1}{\sqrt{2\pi(npq)(\sqrt{\frac{p}{q}}+\frac{x}{\sqrt{n}})(\sqrt{\frac{q}{p}}-\frac{x}{\sqrt{n}})}} &= \frac{1}{\sqrt{2\pi}}\lim_{n\to\infty}\frac{1}{\sqrt{(\sqrt{\frac{p}{q}}+\frac{x}{\sqrt{n}})(\sqrt{\frac{q}{p}}-\frac{x}{\sqrt{n}})}} \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\sqrt{\frac{p}{q}}\sqrt{\frac{q}{p}}}} \\
&= \frac{1}{\sqrt{2\pi}}
\end{aligned}$$

With regard to the second term, we transform each of the two factors using the identity $z = e^{\ln z}$, so that

$$\left(1+\frac{x}{\sqrt{n}}\sqrt{\frac{q}{p}}\right)^{-np-x\sqrt{npq}} = e^{(-np-x\sqrt{npq})\ln\left(1+\frac{x}{\sqrt{n}}\sqrt{\frac{q}{p}}\right)}$$

and resort to the Taylor expansion

$$\ln(1+z) = z - \frac{z^2}{2} + O(z^3)$$

for $z = x\sqrt{q/np}$, where $O(z^3)$ means a sum of terms in z each of degree no less than 3. Therefore we get the exponent of e

$$(-np - x\sqrt{npq}) \ln\left(1 + \frac{x}{\sqrt{n}}\sqrt{\frac{q}{p}}\right) = (-np - x\sqrt{npq})\left(x\sqrt{\frac{q}{np}} - \frac{x^2}{2}\frac{q}{np} + O(x^3\sqrt{\frac{q}{np}}^3)\right)$$

After handling the second factor in the same fashion, we obtain the overall exponent

$$(-np - x\sqrt{npq})\left(x\sqrt{\frac{q}{np}} - \frac{x^2}{2}\frac{q}{np} + O(x^3\sqrt{\frac{q}{np}}^3)\right) + (-nq + x\sqrt{npq})\left(-x\sqrt{\frac{p}{nq}} - \frac{x^2}{2}\frac{p}{nq} + O(x^3\sqrt{\frac{p}{nq}}^3)\right)$$

We now develop the products and appropriately associate the terms, thereby obtaining:

$$\begin{aligned} & (-x\sqrt{npq} + x\sqrt{npq}) + (\frac{x^2 q}{2} + \frac{x^2 p}{2}) - (x^2 p + x^2 q) + P(\frac{1}{\sqrt{n}}) \\ = & -\frac{x^2}{2} + P(\frac{1}{\sqrt{n}}) \end{aligned}$$

where $P()$ is a sum of terms each containing $1/\sqrt{n}$ as a factor, so that $\lim_{n\to\infty} P(1/\sqrt{n}) = 0$. We conclude therefore that

$$\lim_{n\to\infty} \sqrt{npq} B(n,p;j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$