Chapter 11

Convergence theorems

We've already discussed the difficulty in defining the probability measure in terms of an experimental frequency measurement. The heart of the problem lies in the definition of the limit, and this was set aside in favour of the axiomatic basis.

Even, within the axiomatic premise, we have several important convergence theorems relating to a large number of repeated trials (sampling). The two fundamental theorems are:

- the law of large numbers
- the central limit theorem

These convergence theorems are the foundations of statistics. We can consider the subject of Statistics as *experimental probability*. That is the collection and analysis of data (measurements) to determine properties of the underlying theory (probability). These relations underpin that subject in that they support that the measurement process (under certain conditions) will give progressively better estimates.

11.1 Sampling

To simplify the discussion, suppose we have a discrete random variable, X. This random variable has a known (theoretical) mean and standard deviation defined by:

$$\mu \equiv \mathbb{E}\,(X) \qquad , \qquad \sigma = \sqrt{\mathrm{var}\,(X)} \qquad . \tag{11.0}$$

and calculated from the (known) probability mass function. We will refer to these theoretical quantities as the *true* mean and *true* standard deviation.

Now consider an experiment in which this random variable is sampled n times under identical conditions. The aim of this experimental sampling might be to *infer* or deduce properties of this *theoretical* distribution.

Suppose the experiment involves n samples of the random variable. For example, n rolls of a die, or n individuals chosen from a population. Let us denote the (random) value of X obtained from the ith sample as, X_i. Then one can gather these outcomes together in the following way. Take the *sample sum* of these variables:

$$S_n \equiv X_1 + X_2 + \cdots + X_i + \cdots + X_n \qquad . \tag{11.0}$$

We try to ensure that the individual samples X_i are independent. If the sample is done correctly (and we won't go into what that means here), these values will be independent and identically distributed.

If this is the case, then they will have the same common mean (μ) and variance (σ^2). Then it follows that:

$$\mathbb{E}\,(S_n) = \mathbb{E}\,(X_1 + X_2 + \cdots + X_i + \cdots + X_n) = n\mathbb{E}\,(X_i) = n\mu \quad , \tag{11.0}$$

and that:

$$\mathrm{var}\,(S_n) = \mathrm{var}\,(X_1 + X_2 + \cdots + X_i + \cdots + X_n) = n.\mathrm{var}\,(X_i) = n\sigma^2 \qquad . \tag{11.0}$$

11.2 The *weak law* of large numbers

Let us simply state the law as a theorem and then present a proof.

$$\boxed{\lim_{n\to\infty} P\left(\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right)=0} \quad . \tag{11.0}$$

for *any* $\varepsilon > 0$.

Proof:

Firstly, we know that the *theoretical* expected value of the *sample mean* is the *true mean*, μ, namely:

$$\mathbb{E}\left(\frac{S_n}{n}\right)=\mu \tag{11.0}$$

and that:

$$\mathrm{var}\left(\frac{S_n}{n}\right)=\frac{\sigma^2}{n} \quad . \tag{11.0}$$

According to the Chebyshev inequality (10.40) we have:

$$P\left(\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right)\le\frac{\sigma^2/n}{\varepsilon^2} \quad . \tag{11.0}$$

Then, for any finite ε, taking the limit of an infinitely large sample,$n\to\infty$, we have:

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right)=0 \quad . \tag{11.0}$$

The *meaning* of the theorem is that if we take a sufficiently large sample ($n\to\infty$), then the *sample average* (experiment) converges with certainty to the *true* mean (theory). That is, the probability of the sample average S_n/n (that is the experiment) deviating from μ (the theory) by a given amount ε (*no matter how small*) is zero, in the limit of an infinitely large sample.

This elucidates the importance of the theorem and moreover quantifies the sampling error. So the larger the sample size, the more accurate the sample mean approximates the true mean. Such a conclusion validates any experimental measurements or observations of random processes. Of course, there is a practical problem to enforcing the power of the theorem, the ability to do such experiments is severely limited by intrinsic *sampling* errors (bias) and *measurement* errors. These are broadly lumped together as *statistical errors*.

There is a more impressive version of this law: *the strong law of large numbers*, which has a subtle but important difference:

$$\boxed{P\left(\lim_{n\to\infty}\left|\frac{S_n}{n}-\mu\right|>\varepsilon\right)=0} \quad . \tag{11.0}$$

for any $\varepsilon > 0$. The reason it is called the strong law is that it imposes convergence in a much stronger sense. The weak law refers to the limit of the probabilities, while the strong law describes the probability of the limit. The proof of the strong law is more complicated - that's the price you pay for having a better law. Studying the details of the proof are worth the effort. but will not be covered here. Instead we progress to the second topic, the *central limit theorem*.

11.3 Central-limit theorem

As before, the theorem will be stated and then proven.

Let $\{X_1, X_2, \ldots X_n\}$ be a set of independent, identically-distributed variables, with $\mathbb{E}(X)=\mu$ and $\mathrm{var}(X)=\sigma^2$. Then, with $S_n\equiv X_1+X_2+\cdots+X_n$, we have:

$$\boxed{\lim_{n\to\infty} P\left(\frac{(S_n/n)-\mu}{\sigma/\sqrt{n}}\le z\right)=\Phi(z)} \quad . \tag{11.0}$$

where $\Phi(z)$ is the standard normal probability distribution function:

$$\Phi(z) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}x^2} dx \qquad . \tag{11.0}$$

Proof

What we aim to show is that the sample variables have normal distributions as the sample size increase, $n \to \infty$. We will do this indirectly, by showing that the *moment-generating function* of a sample variable tends to the moment-generating function of a normal distribution. Since there is a unique correspondence between the distribution and its moment-generating functions, this convergence implies that the underlying distributions also converge.

Firstly, recall that the standard normal probability density is given by:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad . \tag{11.0}$$

with $\mathbb{E}(X) = 0$ and $\mathrm{var}(X) = 1$. Then the corresponding *moment-generating function* is:

$$M_X(t) = \mathbb{E}\left(e^{tX}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx - \frac{1}{2}x^2} dx \qquad . \tag{11.0}$$

This integral can be evaluated by completing the square:

$$tx - \tfrac{1}{2}x^2 = \tfrac{1}{2}t^2 - \tfrac{1}{2}(x-t)^2 \quad ,$$

and then changing variable $x \to x - t$, it follows that the moment-generating function for the standard normal distribution is:

$$\boxed{M_X(t) = \mathbb{E}\left(e^{tX}\right) = e^{+\frac{1}{2}t^2}} \qquad . \tag{11.0}$$

Now turning our attention to the *experiment*, given that:

$$S_n \equiv X_1 + X_2 + \cdots + X_n \qquad ,$$

set:

$$Z_n \equiv \frac{(S_n/n) - \mu}{\sigma/\sqrt{n}} \qquad . \tag{11.0}$$

By this transformation, the sample sum S_n is now converted to Z_n, a variable with mean and variance:

$$\mathbb{E}(Z_n) = 0 \qquad , \qquad \mathrm{var}(Z_n) = 1 \qquad . \tag{11.0}$$

The aim is to show that, in the limit $n \to \infty$, not only does this variable have the same mean and variance as the standard normal distribution, but it is *identical* to the standard normal distribution.

We have previously mentioned that all we require for this equality is that their moment-generating functions are the same. So let us consider the moment-generating function of Z_n. By definition we have:

$$\begin{aligned} M_{Z_n}(t) = \mathbb{E}\left(e^{tX}\right) &= \mathbb{E}\left(\exp\left(\frac{t}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu)\right)\right) && (11.1) \\ &= \mathbb{E}\left(\exp\left(\frac{t}{\sigma\sqrt{n}}(X_1 - \mu)\right) \times \cdots \times \exp\left(\frac{t}{\sigma\sqrt{n}}(X_n - \mu)\right)\right) \\ & && (11.1) \end{aligned}$$

And since $\{X_1, \ldots, X_n\}$ are independent variables we can write:

$$M_{Z_n}(t) = \mathbb{E}\left(\exp\left(\frac{t}{\sigma\sqrt{n}}(X_1 - \mu)\right)\right) \times \cdots \times \mathbb{E}\left(\exp\left(\frac{t}{\sigma\sqrt{n}}(X_n - \mu)\right)\right) \qquad . \tag{11.1}$$

Given that the X_i are identically distributed, we have:

$$M_{Z_n}(t) = \left[\mathbb{E}\left(\exp\left(\frac{t}{\sigma\sqrt{n}}(X - \mu)\right)\right)\right]^n \qquad . \tag{11.1}$$

So the moment generating function depends on t. One can make a Maclaurin expansion for the exponential function in t as follows:

$$\begin{aligned}\mathbb{E}\left(\exp\left(\frac{t}{\sigma\sqrt{n}}(X-\mu)\right)\right) &= \mathbb{E}(1)+\mathbb{E}\left(\frac{t}{\sigma\sqrt{n}}(X-\mu)\right)+\mathbb{E}\left(\frac{1}{2!}\frac{t^2}{\sigma^2 n}(X-\mu)^2\right)\\ &+\mathbb{E}\left(\frac{1}{3!}\frac{t^3}{\sigma^3 n^{3/2}}(X-\mu)^3\right)+\cdots \quad .\end{aligned}$$

Now evaluating the expections of each term in turn, and noting that $\mathbb{E}(X)=\mu$ by definition, this gives rise to:

$$\mathbb{E}\left(\exp\left(\frac{t}{\sigma\sqrt{n}}(X-\mu)\right)\right)=1+0+\frac{1}{2!}\frac{t^2}{\sigma^2 n}\sigma^2+\frac{1}{3!}\frac{t^3}{\sigma^3 n^{3/2}}\mathbb{E}\left((X-\mu)^3\right)+\cdots \quad . \tag{11.-2}$$

this finally gives:

$$\mathbb{E}\left(\exp\left(\frac{t}{\sigma\sqrt{n}}(X-\mu)\right)\right)=1+\frac{t^2}{2n}+O\left(\frac{t^3}{n^{3/2}}\right) \quad . \tag{11.-2}$$

Note, the symbol O, in the last term means 'of the order'. That is this term has some finite (unspecified) coefficient but is proportional to $t^3 n^{-3/2}$. At this juncture we take the limit $n \to \infty$, which leads to the result:

$$\lim_{n\to\infty} M_{Z_n}(t)=\lim_{n\to\infty}\left[\mathbb{E}\left(\exp\left(\frac{t}{\sigma\sqrt{n}}(X-\mu)\right)\right)\right]^n=\lim_{n\to\infty}\left[1+\frac{t^2}{2n}+O\left(\frac{t^3}{n^{3/2}}\right)\right]^n \tag{11.-2}$$

This culminates with the following expression for the moment-generating function.

$$\lim_{n\to\infty} M_{Z_n}(t)=e^{\frac{1}{2}t^2} \quad . \tag{11.-2}$$

That is the variable Z_n has the same moment generating function as the *standard normal distribution*: equation 11.3. Consequently the probability distributions are identical, in every aspect, in this limit.

end of proof.

In summary, the distribution of the *sample* of *any* discrete random variable, tends towards a normal distribution.

Consider how this works in practice. For any discrete random variable X with mean, μ, and (finite) standard deviation, σ, the sample (sum) variable:

$$S_n \equiv X_1+\cdots+X_n \quad , \tag{11.-2}$$

obeys the following relation, for large n,

$$P\left(\frac{(S_n/n)-\mu}{\sigma/\sqrt{n}}\le z\right)\approx\Phi(z) \quad . \tag{11.-2}$$

Equivalently, rearranging this we can find the probability distribution of S_n:

$$F_{S_n}(s)=P(S_n\le s)=\Phi\left(\frac{s/n-\mu}{\sigma/\sqrt{n}}\right) \quad . \tag{11.-2}$$

While S_n is, of course a *discrete* variable, according to the central-limit theorem, for a large sample it is can be approximated by a *continuous* normal distribution.

Consequently the probability density will have the form:

$$f_{S_n}(s)=\frac{d}{ds}F_{S_n}(s)=\frac{1}{\sigma\sqrt{n}}\phi\left(\frac{s/n-\mu}{\sigma/\sqrt{n}}\right) \quad , \tag{11.-2}$$

that is:

$$\boxed{f_{S_n}(s)=\frac{1}{\sigma\sqrt{n}\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{S_n/n-\mu}{\sigma/\sqrt{n}}\right)^2\right]} \quad . \tag{11.-2}$$

This is a continuous variable, and to convert this to the equivalent *probability mass* for the discrete variable S_n, we use the fact that, by definition

$$\boxed{P(s \le S_n \le s + \Delta s) = f_{S_n}(s)\Delta s} \quad , \tag{11.-2}$$

and when the interval is , $\Delta s = 1$, as it would for a counting variable, for example the number of HEADS in n tosses of a coin, we would get:

$$P(S_n = s) = f_{S_n}(s) = \frac{1}{\sigma\sqrt{n}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{S_n/n - \mu}{\sigma/\sqrt{n}}\right)^2\right] \quad , \quad S_n = 0, 1, 2, \ldots, n \tag{11.-2}$$

for the *probability mass function* for the discrete variable S_n

EXAMPLE

A certain flight carries a random number of passengers, X, such that the average and variance are given by:

$$\mathbb{E}(X) = 50 \quad , \quad \text{var}(X) = 100 \quad . \tag{11.-2}$$

In a sample of 20 flights:

(a) What is the probability that the total number of passengers is less than 950 ?

(b) Calculate the probability that the passenger total is *exactly* 1010.

SOLUTION

Although $n = 20$ is not a very large number, let's use the central-limit theorem anyway. Then the total number of passngers (over all 20 flights) we will call:

$$S_n = X_1 + X_2 + \cdots + X_n \quad . \tag{11.-2}$$

Then we have the correspondence:

$$n = 20 \qquad \mu = 50 \qquad \sigma = \sqrt{100} = 10 \quad . \tag{11.-2}$$

Then to answer (a) we are after

$$P(S_n \le S) \qquad \text{with} \qquad S = 950 \quad . \tag{11.-2}$$

This is equal to, making the same changes to both sides of the inequality:

$$P(S_n \le S) = P(S_n/n - \mu \le S/n - \mu) = P\left(\frac{S_n/n - \mu}{\sigma/\sqrt{n}} \le \frac{S/n - \mu}{\sigma/\sqrt{n}}\right) \tag{11.-2}$$

And, according to the central-limit theorem:

$$P(S_n \le S) \approx P\left(\frac{S_n/n - \mu}{\sigma/\sqrt{n}} \le \frac{950/20 - 50}{10/\sqrt{10}}\right) \approx \Phi\left(\frac{950/20 - 50}{10/\sqrt{10}}\right) = \Phi(-1.118) \tag{11.-2}$$

Now, due to the reflection symmetry of the standard normal distribution (an even function) we have the following identities

$$\Phi(-z) = 1 - \Phi(z) \tag{11.-2}$$

and for this reason, the published tables only need to provide values for $z \ge 0$. From the tables:

$$P(S_n \le 950) = 1 - \Phi(1.118) \approx 0.132 \quad . \tag{11.-2}$$

This provides the answer to part (a).

To answer (b) we reframe the question as, what is the value of: $P(S_n \le 1010.5) - P(S_n \le 1009.5)$? Then, using the same arguments, we arrive at the answers:

$$P(S_n \le 1010.5) = \Phi(0.2348) = 0.5928 \quad , \quad P(S_n \le 1010.5) = \Phi(0.2124) = 0.5841 \quad . \tag{11.-2}$$

Thus

$$P(S_n = 1010) \approx 0.5928 - 0.5841 = 0.0087 \quad . \tag{11.-2}$$

11.4 Confidence limits and standard error

Suppose you wish to conduct a survey, for example to determine the popularity of the government or the fraction of the student population that smokes. In sampling a finite number n of people there will inevitably be uncertainty (randomness) in the outcome. Nonetheless it is possible (using the central-limit theorem) to estimate the error in your estimates based on the size of the sampling. One can define *confidence limits* as the probability of the correctness of your answer (within a certain accuracy) based on the size of the sample.

Let us set aside, for the present, the formidable challenge of choosing who to sample and how to sample them, and focus on the size of sample required to form an opinion. That is a different kind (and more challenging problem) of additional statistical error. Suppose, you wish to estimate some parameter, let's call it z, with a certain accuracy (error), let's call it ε, within a degree of certainty. Let's call the certainty, or more strictly the probability, p. Statisticians tend to get a bit obsessed with statistical tests and p-values. How many people n, would you need to sample to get a desired accuracy ? Clearly, this will depend in some way on the size of error you are willing to tolerate and the degree of certainty you wish to impose.

A concrete example will illustrate the role of the central-limit theorem in providing such an estimate. We consider a yes/no question and let us say that you are conducting a survey on, let's say the fraction of people who smoke, or the percentage of people who went to the movies in the last month, or the fraction of the population that support the government. Suppose that the true value (that is if we sampled the entire population) of people who say they like classical music is $0 \le z \le 1$. We seek a value for z which is accurate to $\varepsilon = \pm 0.05$ (this is not quite the same as saying a 5% accuracy), and we want (at *least*) $p = 90\%$ for the confidence limit.

One must be careful about the use of the term *confidence* in this context. This does *not* mean we are 90% confident that the estimate is correct. It simply means that 'nine times out of ten' we expect this result. The term *confidence limit* is an unfortunate choice of words in this context.

So, the survey approaches n individuals and, of these, s people say they like classical music. In mathematical terms we wish to know the value n such that the sample fraction, s/n, that said 'yes' (the like classical music) is such that the probability *exceeds* a certain value, p_c (the *confidence limit*):

$$P\left(\left|\frac{s}{n} - z\right| \le \varepsilon\right) \ge p_c \qquad . \qquad (11.\text{-}2)$$

If the answer to the question is simply yes/no (Bernoulli variable) then the theoretical mean would be z and the standard deviation for any individual would be: $\sigma = \sqrt{z(1-z)}$. Thus we seek n such that:

$$P\left(\frac{-\varepsilon}{\sigma/\sqrt{n}} \le \frac{s/n - z}{\sigma/\sqrt{n}} \le \frac{+\varepsilon}{\sigma/\sqrt{n}}\right) \ge p_c$$

For a large sample, we can use the central-limit theorem (normal distribution for the sample average), so that:

$$-1 + 2\Phi\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right) \ge p_c$$

That is one would require a sample size given by:

$$\boxed{n \ge \left(\frac{\sigma}{\varepsilon}\right)^2 \left[\Phi^{-1}\left(\frac{p_c + 1}{2}\right)\right]^2} \qquad . \qquad (11.\text{-}2)$$

For the example discussed above, using tables for the standard normal distribution we find $\Phi^{-1}(0.95) \approx 1.645$, and since $\sigma^2 \le \frac{1}{4}$ we have:

$$n \ge 271 \qquad .$$

So, we should sample at least $n = 271$ people, to have a 90% confidence limit, that our experimental answer to the value of z is within the error 0.05. Thus the sample size, n is a function of the accuracy desired (ε) as well as the confidence required (p_c), and is proportional to the variance (σ^2) of the variable.

By the same token, for a fixed confidence value (p_c), the *error* (ε) in choosing a sample of size n would be:

$$\boxed{|\varepsilon| \ge \frac{\sigma}{\sqrt{n}} \Phi^{-1}\left(\frac{p+1}{2}\right)} \qquad . \qquad (11.\text{-}2)$$

This dependence is known to every statistician, and indeed the expression:

$$\frac{\sigma}{\sqrt{n}} \tag{11.-2}$$

is so widely encountered, it is called *the standard error*. However, one must be careful not to quote this expression as 'the error'. That is, it would be wrong to state that the estimate for z is limited within these bounds. The standard error only has a meaning in terms of the normal distribution.

Commonly, one seeks confidence at the 95% level, in which case one can be more precise: $\Phi^{-1}(0.975) \approx 1.96$ and thus:

$$|\varepsilon| \geq \frac{\sigma}{\sqrt{n}} 1.96 \qquad . \tag{11.-2}$$

That is the right-hand side provides a lower bound on the error.

Therefore, while our sampling error diminishes with certainty as the sample size increases, the rate of decrease in error is frustrating slow, namely $\sim n^{-\frac{1}{2}}$. So making the sample 10 times larger only leads to a factor 3 in reduction in sample error.

11.5 Convergence to the normal distribution

We have used the moment-generating function to prove the central-limit theorem. The power of this approach is that it works for *any* distribution function.

We can show explicitly the convergence to a normal distribution directly from the probability mass. Consider a series of trials (such as coin tossing) where $W_i \in \{0, 1\}$ corresponds to the ith toss of the coin producing a TAILS or HEADS, respectively. Our sample is then a (large) number of coin tosses N, and let us take the sample random variable as the total number of HEADS.

On any given toss the probability of heads is p. Thus

$$X \equiv W_1 + W_2 + \cdots + W_n \qquad . \tag{11.-2}$$

Then clearly, the W_i are independent, so that:

$$\mathbb{E}(X) = np \qquad , \qquad \text{var}(X) = npq \qquad . \tag{11.-2}$$

as is well known for a binomial distribution.

Then the probability mass for X (the sample) is just the binomial distribution

$$f_X(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \qquad .$$

We note one property of this function of x. For $np \gg 1$, the function reaches a single maximum, for a particular value of x and then falls away rapidly as $x \to 0.$ or $x \to \infty$. This is illustrated by the bar chart (figure **??**).

The correspondence with the normal distribution has been known for a very long time, at least as far back as de Moivre in 1733. Let us go through the argument here.

First of all, we need to consider a very large sample size $n \gg 1$. Thus means evaluating large factorials and Stirling's formula (see appendix for derivation) gives us a very good approximation for this, namely that:

$$n! \approx \sqrt{2\pi n}\; n^n e^{-n} \qquad . \tag{11.-2}$$

That is

$$\ln n! \approx \tfrac{1}{2}\ln(2\pi) + (n + \tfrac{1}{2})\ln n - n \qquad . \tag{11.-2}$$

Now, let us define: $g(x) = \ln f_X(x)$ so that, for large n, x and $n - x$ we have:

$$\begin{aligned} g(x) \quad \approx \quad & \left[\tfrac{1}{2}\ln(2\pi) + (n + \tfrac{1}{2})\ln n - n\right] - \left[\tfrac{1}{2}\ln(2\pi) + (x + \tfrac{1}{2})\ln x - x\right] \\ & - \left[\tfrac{1}{2}\ln(2\pi) + (n - x + \tfrac{1}{2})\ln(n - x) - (n - x)\right] \\ & \quad + x\ln p + (n - x)\ln q \qquad . \qquad (11.-3) \\ & \qquad (11.-2) \end{aligned}$$

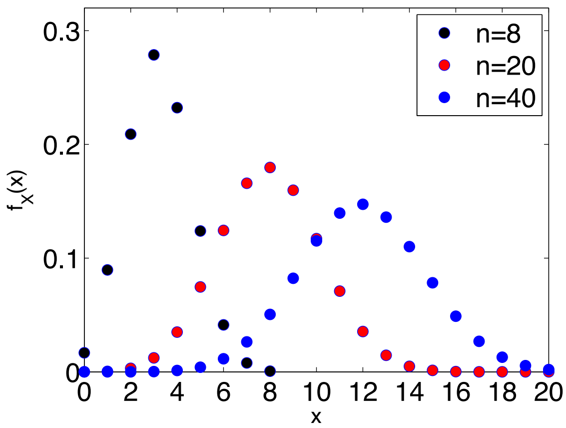

Figure 11.1: The binomial distribution for $p = 0.4$, $n = 8, 20, 40$. Note that the distribution has a single maximum, near $x = np$, and that as n increases, the shape begins to resemble the 'bell-shaped' normal distribution.

Now, as indicated in figure 11.1, $f(x)$ and hence $g(x)$, has a maximum value at some intermediate value of x.

We can find the location of the maximum by finding the solution of:

$$g'(x) = 0 \tag{11.-2}$$

That is:

$$g'(x) = -\ln x - 1 - \frac{\frac{1}{2}}{x} + \ln(n-x) + 1 + \frac{\frac{1}{2}}{(n-x)} + \ln p - \ln q \qquad . \tag{11.-2}$$

Now for large xn, and $n - x$ the two terms in $1/x$ and $1/(n-x)$ can be neglected to a fair approximation. So the value of x where the maximum occurs is approximately the solution of:

$$\ln\left(\frac{(n-x_0)p}{x_0 q}\right) = 0 \tag{11.-2}$$

That is:

$$x_0 = np \qquad . \tag{11.-2}$$

Recall that this value is the mean of the binomial distribution. We now see that this becomes the mode of the sample variable, X.

Now the second derivative at this point, $g''(x_0)$, is given by:

$$g''(x_0) \approx -\frac{1}{x_0} - \frac{1}{n-x_0} = -\frac{1}{npq} \qquad . \tag{11.-2}$$

Clearly $g''(x_0) < 0$, which confirms that the stationary point x_0 is a maximum.

Then consider the Taylor series of $g(x)$ in the neighbourhood of the maximum:

$$g(x) \approx g(x_0) + (x - x_0)g'(x_0) + \tfrac{1}{2}(x - x_0)^2 g''(x_0) \tag{11.-2}$$

Then, since $g'(x_0) = 0$ and,

$$g(x_0) = -\tfrac{1}{2}\ln(2\pi) + (n + \tfrac{1}{2})\ln n + (np + \tfrac{1}{2})\ln np - (nq + \tfrac{1}{2})\ln nq + np\ln p + nq\ln q \qquad . \tag{11.-2}$$

That is, after simplification:

$$g(x_0) = -\tfrac{1}{2}\ln(2\pi) - \tfrac{1}{2}\ln n - \tfrac{1}{2}\ln pq \tag{11.-2}$$

Then:

$$g(x) \approx -\tfrac{1}{2}\ln(2\pi) - \tfrac{1}{2}\ln n - \tfrac{1}{2}\ln pq - \frac{1}{2npq}(x - np)^2 \qquad . \tag{11.-2}$$

It follows that, in the limit of a large sample, we have the approximation:

$$\boxed{f_X(x) \approx \frac{1}{\sqrt{2\pi}\,\sqrt{npq}} \exp\left(-\frac{(x-np)^2}{2npq}\right)} \qquad . \tag{11.-2}$$

So we find that X has a normal distribution with mean given by $\mu = np$ and variance npq.

We see that the result concurs with the central-limit theorem, since for the geometric variable

$$\mu = p \qquad , \qquad \sigma = \sqrt{pq} \tag{11.-2}$$

Then according to the central-limit theorem, in the form (11.3)

$$\boxed{P(S_n = s) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(s-np)^2}{2npq}}} \qquad . \tag{11.-2}$$

which agrees with (11.5).