Homework 3 Clarification + a Bonus

27 October 2023

Imagine an “experiment” that has a binary result, success vs. failure. For example, flipping a coin, success=heads, failure=tails.
I want to estimate the probability \( p \) of success by repeating the experiment many times. In the flipping the coin example, I want to estimate the probability of getting heads (it should be 50% if the coin is “fair”).
I perform the experiment \( N \) times. The number of successes (heads, in the coin example) is \( S \). Then my estimate of \( p \) is \( \hat{p} = S/N \). This is not the ”true” value of \( p \) because there will be random fluctuations in the number of successes \( S \). This is an example of a stochastic (ie: random) binomial process.
Let \( p_0 \) be the true value of \( p \). On average I will get \( \hat{p} = p_0 \), ie, \( S = Np_0 \). But this is only on average. In reality the experimental value \( S \) will fluctuate around the average (or mean). It can be shown that as long as \( N \) is large enough, and \( p_0 \) is not too close to zero or one, \( S \) will be distributed approximately according to a Gaussian ("bell curve") of mean \( Np_0 \) and standard deviation \( \sigma(S) = \sqrt{Np_0(1-p_0)} \).
Since \( \hat{p} = S/N \), and \( S \) has standard deviation \( \sigma(S) \), my estimate \( \hat{p} \) will have a standard deviation \( \sigma(p) = \sigma(S)/N = \sqrt{p_0(1-p_0)/N} \). The (relative) accuracy or precision is then \( \sigma(p)/p = (\sqrt{p_0(1-p_0)/N})/p_0 \).
Note that as \( N \) increases, my estimate of \( p \) gets better and better (\( N \) is in the denominator). This makes sense and it is to be expected. What is perhaps not so obvious is that the precision goes like the \( 1/\sqrt{N} \).

1 Bonus part: derivation of statements mentioned above

1.1 Probability of observing \( S \) out of \( N \) for a given \( p_0 \)
Assume we have \( S \) successes out of \( N \). Think of a series of \( N \) trials, with \( S \) successes (denoted by 1) and \( N - S \) failures denoted by 0. A particular such series might look like this:

\[
1,1,0,0,0,1,1,\ldots,0,0,1
\]

Since there are \( S \) successes and \( N - S \) failures, such a series will have probability \( p_0^S(1-p_0)^{N-S} \). But there are \( \binom{N}{S} \) equally likely ways of rearranging the 1’s and
0's to obtain different series with \( S \) successes. Therefore the probability of \( S \) successes out of \( N \) trials is

\[
p(S \text{ out of } N) = \binom{N}{S} p_0^S (1 - p_0)^{N-S} = \frac{N!}{S!(N-S)!} p_0^S (1 - p_0)^{N-S}
\]

### 1.2 Mean and Standard Deviation of \( S \), the easy way

Notation:

- \(< A > = \text{mean of } A\)
- \(\text{Var}(A) = < A^2 > - < A >^2 = \text{variance of } A = \sigma^2(A)\). The standard deviation of \( A \) is the square root of the variance, ie, \( \sigma(A) \)

We use the trick that \(< x + y >= < x > + < y > \) and \(\text{Var}(x + y) = \text{Var}(x) + \text{Var}(y)\) where \( x \) and \( y \) are independent random variables. We take the \( N \) trials as \( N \) independent random variables \( x_i \) with \( x_i = 0 \) or \( 1 \) for failure or success, so that \(< S >= N \cdot < x_i > \) and \(\text{Var}(S) = N \cdot \text{Var}(x_i)\). Then

\[
< x_i >= 1 \cdot p_0 + 0 \cdot (1 - p_0) = p_0 \quad \rightarrow \quad < S >= Np
\]

(not a surprise...). For the variance:

\[
< x_i^2 >= 1^2 \cdot p_0 + 0^2 \cdot (1 - p_0) = p_0
\]

\[
\text{Var}(x_i) = < x_i^2 > - < x_i >^2 = p_0 - p_0^2 = p_0(1 - p_0) \quad \rightarrow \quad \sigma^2(S) = Np_0(1 - p_0)
\]

or

\[
\sigma(S) = \sqrt{Np_0(1 - p_0)}
\]

### 1.3 Mean of \( S \), the hard way

The mean of \( S \) is given by

\[
< S >= \sum_{I=0}^{I=N} I \cdot p(I \text{ out of } N) = \sum_{I=0}^{I=N} I \binom{N}{I} p_0^I (1-p_0)^{N-I} = \sum_{I=0}^{I=N} I \frac{N!}{I!(N-I)!} p_0^I (1-p_0)^{N-I}
\]

Since the first term in the sum is zero, I can just start the sum from \( I = 1 \) instead of \( I = 0 \)

\[
< S >= \sum_{I=1}^{I=N} I \frac{N!}{I!(N-I)!} p_0^I (1-p_0)^{N-I}
\]
We now use the identity $I \binom{N}{i} = N \binom{N-1}{i-1}$, which gives:

$$< S > = \sum_{i=1}^{I=N} N \binom{N-1}{i-1} p_i^0 (1 - p_0)^{N-I} = N \sum_{i=1}^{I=N} \binom{N-1}{i-1} p_i^0 (1 - p_0)^{N-I}$$

Now we write $p_i^I = p_0 p_i^{I-1}$:

$$< S > = N p_0 \sum_{i=1}^{I=N} \binom{N-1}{i-1} p_i^{I-1} (1 - p_0)^{N-I}$$

Which I can rewrite as

$$< S > = N p_0 \sum_{i=1}^{I=N} \binom{N-1}{i-1} p_i^{I-1} (1 - p_0)^{(N-1)-(I-1)}$$

Write $J = I - 1$ and $M = N - 1$:

$$< S > = N p_0 \sum_{J=0}^{J=M} \binom{M}{J} p_J^0 (1 - p_0)^{M-J}$$

$$< S > = N p_0 \sum_{J=0}^{J=M} p(J \text{ out of } M)$$

But the sum of $p(J \text{ out of } M)$ over all possible values of $J$ is one. Therefore:

$$< S > = N p_0$$

### 1.4 Standard deviation of $S$, the hard way

We start by calculating the mean of $S^2$. Proceeding as before

$$< S^2 > = \sum_{I=0}^{I=N} I^2 \ p(I \text{ out of } N) = \sum_{I=0}^{I=N} I^2 \ \binom{N}{I} p_I^0 (1 - p_0)^{N-I}$$

As before, we start the sum at $I = 1$, use the identity $I \binom{N}{1} = N \binom{N-1}{1}$. write $p_I^I = p_0 p_I^{I-1}$, $J = I - 1$ and $M = N - 1$. We arrive at:

$$< S^2 > = N p_0 \sum_{J=0}^{J=M} (J + 1) \ \binom{M}{J} p_J^0 (1 - p_0)^{M-J}$$

Let’s split the sum into two pieces:

$$< S^2 > = N p_0 \sum_{J=0}^{J=M} p_J^0 (1 - p_0)^{M-J} + N p_0 \sum_{J=0}^{J=M} J \ p_J^0 (1 - p_0)^{M-J}$$
\begin{equation}
<S^2> = Np_0 \sum_{J=0}^{J=M} p(J \text{ out of } M) + Np_0 \sum_{J=0}^{J=M} J \ p(J \text{ out of } M)
\end{equation}

The first sum is equal to one (just as we argued at the end of Section 1.3). The second sum is the mean of \( M \) trials, which is \( Mp_0 = (N-1)p_0 \). Thus:
\begin{equation}
<S^2> = Np_0 + N(N-1)p_0^2 = N^2p_0^2 + Np_0(1-p_0)
\end{equation}

The variance of \( S \) is
\begin{equation}
\sigma^2(S) = <S^2> - <S>^2 = N^2p_0^2 + Np_0(1-p_0) - (Np_0)^2 = Np_0(1-p_0)
\end{equation}

The standard deviation is the square root of the variance, thus
\begin{equation}
\sigma(S) = \sqrt{Np_0(1-p_0)}
\end{equation}

### 1.5 Gaussian Approximation

Here we derive a Gaussian approximation valid for large \( N \) and \( p_0 \) not too close to 1 or zero. The approximation is valid in the “neighborhood” of the mean \(<S> = Np_0\). Note since \( N \) is large and \( p_0 \) is not close to 0 or 1, in the neighborhood of the mean both \( S \) and \( N - S \) are also large.

We start with the equation for \( p(S \text{ out of } N) \) from Section 1.1, and apply Stirling’s approximation for the factorial\footnote{https://tinyurl.com/ywtdna9d}:
\begin{equation}
k! \approx k^k e^{-k} \sqrt{2\pi k}
\end{equation}

(this approximation works quite well, and gets better and better as \( k \) gets larger and larger). Then, after some boring algebra:
\begin{equation}
p(S \text{ out of } N) \approx \left(\frac{Np}{S}\right)^s \left(\frac{Nq}{N-S}\right)^{N-s} \sqrt{\frac{N}{2\pi S(N-S)}}
\end{equation}

where for simplicity I wrote \( p = p_0 \) and \( q = (1-p_0) \). For reasons that will become clear later, I rewrite
\begin{equation}
p(S \text{ out of } N) \approx Be^{\log A}
\end{equation}

where
\begin{equation}
B = \sqrt{\frac{N}{2\pi S(N-S)}}
\end{equation}

and
\[
\log A = \log \left( \frac{N_p}{S} \right)^S \left( \frac{N_q}{N - S} \right)^{N - S}
\]

Now I define \( \delta = S - N_p \). Note that \( N_p \) is the mean, so \( \delta \) is the deviation from the mean. Since we are in the neighborhood of the mean, \( \delta \ll N_p \). Now consider the quantity

\[
\log \frac{N_p}{S} = \frac{N_p}{\delta + N_p} = -\log(1 + \frac{\delta}{N_p})
\]

Similarly:

\[
\log \frac{N_q}{N - S} = \log \frac{N_q}{N - \delta} = -\log(1 - \frac{\delta}{N_q})
\]

Another key approximation is based on the fact that \( \frac{\delta}{N_p} \ll 1 \) and \( \frac{\delta}{N_q} \ll 1 \). This then allows us to use the approximation \( \log(1 + x) \approx x - \frac{1}{2}x^2 \), valid for small \( x \) and good up to terms of order \( x^3 \). Using this expansion and also the relationships \( \delta = S - N_p \) and \( N - S = N_q - \delta \):

\[
\log A \approx -\left( \frac{\delta + N_p}{N_p} \right) \left( \frac{\delta}{N_p} - \frac{\delta^2}{2N^2p^2} \right) - \left( \frac{N_q - \delta}{N_q} \right) \left( -\frac{\delta}{N_q} - \frac{\delta^2}{2N^2q^2} \right)
\]

Multiplying through and dropping terms proportional to \( \delta^3 \):

\[
\log A \approx -\frac{\delta^2}{2Np} - \delta - \frac{\delta^2}{2Nq} + \delta = -\frac{\delta^2}{2Npq}(p + q) = -\frac{\delta^2}{2Npq}
\]

since \( p + q = 1 \). Going back to the equation for \( B \):

\[
B = \sqrt{\frac{N}{2\pi S(N - S)}} = \sqrt{\frac{N}{2\pi(\delta + N_p)(N_q - \delta)}} = \sqrt{\frac{N}{2\pi(N^2pq + Nq\delta - Np\delta - \delta^2)}}
\]

Since \( \delta \) is small, we can drop the terms proportional to \( \delta \) in the denominator and write

\[
B \approx \sqrt{\frac{1}{2\pi Npq}}
\]

Now that we have approximate expressions for \( B \) and \( \log A \), we can plug them back into the original equation for \( p(S \text{ out of } N) \):

\[
p(S \text{ out of } N) \approx B e^{\log A} \approx \sqrt{\frac{1}{2\pi Npq}} \exp \left( -\frac{\delta^2}{2Npq} \right)
\]

But \( \delta = S - N_p = S - < S > \), where \( < S > \) was the mean obtained in Section 1.2; also, in Section 1.2 we had the variance \( \sigma^2(S) = Npq = Np_0(1 - p_0) \). Thus
\[ p(S \text{ out of } N) \approx \sqrt{\frac{1}{2\pi\sigma^2(S)}} \exp\left(-\frac{(S-\langle S \rangle)^2}{2\sigma^2(S)}\right) = \frac{1}{\sqrt{2\pi\sigma(S)}} \exp\left(-\frac{(S-\langle S \rangle)^2}{2\sigma^2(S)}\right) \]

which is the equation of Gaussian of mean \( \langle S \rangle \) and standard deviation \( \sigma(S) \).