Introduction to Probability and Statistics

Justifications of some formulas

March 19, 2009

We have encountered a number of important formulas in class recently, that we did not give justifications for. Here are some sketches that might give you a feel for why these are true.

1 The Central Limit Theorem

The Central Limit Theorem says that if $X_1, \ldots, X_n$ are independent random variables, all with the same distribution that has mean μ and variance σ^2, then for large n

$$X_1 + \ldots + X_n \approx \mathcal{N}\left(n\mu, n\sigma^2\right).$$

In other words, the sum of independent copies of the same random variable has approximately a normal distribution. Scaling to turn the right hand side into a standard normal, there is a more precise statement:

$$P\left(\frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \leq z\right) \to \Phi(z) = P(Z \leq z) \quad \text{as } n \to \infty$$

where Z is a standard normal.

One way to justify this is to write

$$\begin{aligned} \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} &= \left(\frac{X_1 - \mu}{\sigma\sqrt{n}}\right) + \ldots + \left(\frac{X_1 - \mu}{\sigma\sqrt{n}}\right) \\ &= Y_1 + \ldots + Y_n. \end{aligned}$$

Each of the Y_i's have mean 0 and variance $\frac{1}{n}$. That means that they have $E(Y_i) = 0$ and $E(Y_i^2) = \frac{1}{n}$, since $Var(Y_i) = E(Y_i^2) - (E(Y_i))^2$. This tells us something about the moment generating function of Y_i. If the moment generating function of Y_i begins

$$\phi_{Y_i}(t) = a + bt + ct^2 + \ldots$$

then since $\phi_{Y_i}(0) = 1$ we must have $a = 1$; $\phi'_{Y_i}(0) = E(Y_i) = 0$ we must have $b = 0$; and since $\phi''_{Y_i}(0) = E(Y_i^2) = \frac{1}{n}$ we must have $c = \frac{1}{2n}$. So

$$\phi_{Y_i}(t) = 1 + \frac{t^2}{2n} + \text{terms involving } t^3 \text{ and higher powers.}$$

What's the moment generating function of $Y_1 + \ldots + Y_n$? It's

$$\begin{aligned}
\phi_{Y_1+\ldots+Y_n}(t) &= E\left(e^{t(Y_1+\ldots+Y_n)}\right) \\
&= E\left(e^{tY_1}\ldots e^{tY_n}\right) \\
&= E\left(e^{tY_1}\right)\ldots E\left(e^{tY_n}\right) \\
&= \phi_{Y_1}(t)\ldots\phi_{Y_n}(t) \\
&= \left(1+\frac{t^2}{2n} + \text{terms involving } t^3 \text{ and higher powers}\right)^n.
\end{aligned}$$

In the third line above we used the fact that the Y_i's are independent.

Now we have to remember some calculus. One way in which the exponential function e^x is often defined is by

$$e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n.$$

(Reality check: pick a value for x, say $x = 2$, and on a calculator compute

$$\left(1+\frac{2}{1}\right)^1, \left(1+\frac{2}{10}\right)^{10}, \left(1+\frac{2}{100}\right)^{100} \text{ and } \left(1+\frac{2}{1000}\right)^{1000},$$

and see how the values compare to e^2.)

Here's how this is useful: at least for small values of t,

$$\left(1+\frac{t^2}{2n} + \text{terms involving } t^3 \text{ and higher powers}\right)^n \approx \left(1+\frac{t^2}{2n}\right)^n$$

and so by the definition of e^x it's approximately $e^{t^2/2}$. So, at least for small t,

$$\phi_{Y_1+\ldots+Y_n}(t) \approx e^{\frac{t^2}{2}}.$$

But $e^{t^2/2}$ is the moment generating function of the standard normal random variable Z, as we derived in class. This strongly suggests that

$$Y_1 + \ldots + Y_n \approx Z,$$

which is our rough form of the Central Limit Theorem. (To make this precise takes, unfortunately, a lot of blood and sweat!)

2 The formula for sample variance

If $X_1, \ldots, X_n$ is a random sample from a population with mean μ and variance σ^2, then we defined its sample variance to be

$$S^2 = \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}{n-1}$$

where $\bar{X} = \frac{X_1+\ldots+X_n}{n}$ is the sample mean. The reason we divide by $n-1$ and not n is because we want S^2 to be a good estimator for σ^2 in the sense that "on average" it's right; that is, we want $E(S^2) = \sigma^2$. Here's the calculation that shows that dividing by $n-1$ is exactly the right thing to achieve this:

$$\begin{aligned}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 &= \sum_{i=1}^{n}\left(X_i-\frac{X_1+\ldots+X_n}{n}\right)^2\\ &= \sum_{i=1}^{n}\left(\frac{nX_i-(X_1+\ldots+X_n)}{n}\right)^2\\ &= \sum_{i=1}^{n}\left(\frac{-X_1-X_2-\ldots+(n-1)X_i-X_{i+1}-\ldots-X_n}{n}\right)^2\\ &= \frac{1}{n^2}\sum_{i=1}^{n}\left(-X_1-X_2-\ldots+(n-1)X_i-X_{i+1}-\ldots-X_n\right)^2.\end{aligned}$$

When we square each of the n terms inside the sum, and add them, we get X_1^2 a total of $(n-1)^2 + (n-1) = n(n-1)$ times: $(n-1)^2$ from the first term, which has an $(n-1)X_1$ in it, and once from each of the remaining terms (which all have $-X_1$ in them). Similarly we get X_i^2 a total of $n(n-1)$ times for each of the other i's. We get X_1X_2 a total of $-2(n-1) - 2(n-1) + 2(n-2) = -2n$ times: $-2(n-1)$ from the first term, which has an $(n-1)X_1 - X_2$ in it, $-2(n-1)$ from the second term, which has a $-X_1 + (n-1)X_2$ in it, and -2 times from each of the remaining $n-2$ terms (which all have $-X_1 - X_2$ in them). Similarly we get X_iX_j a total of $-2n$ times for each of the other choice of i, j. So

$$\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 = \sum_{i=1}^{n} n(n-1)X_i^2 + \sum_{i\neq j} -2nX_iX_j.$$

When we take the expectation of both sides, every time we encounter an $E(X_i^2)$ we can write $E(X_1^2)$ (since these are the same), and every time we encounter an $E(X_iX_j)$ we can write $E(X_1)^2$ (since by independence $E(X_iX_j) = E(X_i)E(X_j) = E(X_1)^2$). How many times do we encounter an $E(X_i^2)$? $n(n-1)$ times for each i, so $n^2(n-1)$ times in all. How many times do we encounter an $E(X_iX_j)$? $-2n$ times for each $i \neq j$, and there are $\binom{n}{2}$ choices for $i \neq j$, so $-2n\binom{n}{2} = -n^2(n-1)$ times in all. So we get

$$\begin{aligned}E\left(\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right) &= \frac{1}{n^2}\left(n^2(n-1)E(X_1^2) - n^2(n-1)E(X_1)^2\right)\\ &= (n-1)\left(E(X_1^2)-E(X_1)^2\right)\\ &= (n-1)Var(X_1)\end{aligned}$$

and so

$$E(S^2) = \sigma^2.$$