Lecture 12. The Poisson Random Variable

How many chocolate chips?.

What is the probability of getting k chocolate chips in a cookie if the average number of chips per cookie is λ?

You are to assume that the cookies have been cut out from an huge batch of cookie dough in which the chocolate chips are randomly distributed. Actually, we want to find the limit as the amount of dough approaches infinity. (Those of you who like chocolate-chip-cookies should enjoy this.) It turns out that the probability is given by a famous pmf, called the *Poisson distribution*. We shall derive the formula for it from the binomial distribution.

Setting up the problem.

Let us begin by deciding to measure the dough in units of weight such that one cookie weighs one unit. Assume the amount of dough is d—that is, we have enough dough to make d cookies. Let us also assume that the total number of chips in the dough is λd.

Now we will do something that may seem peculiar—or at least picky and obsessive. Pick out one chip from among all of them, and pick out one cookie from all those made. We want to analyze the probability that this one special designated chip is in this one special designated cookie. But this is easy. The probability is $1/d$, since we could make d cookies from the whole batch, and the special chip we picked has an equal chance of being in any one of them. It turn out that this picky, obsessive observation—combined with some mathematical machinery that we already know well—provides the solution of the problem.

When we grab the dough for a single cookie, we may view each chip as being on trial. A chip wins if it gets into the cookie, and it looses if its left behind. Since all the chips are on trial, we have λd trials. Therefore, the number of chips in our cookie is given by the random variable $X =$ the number of successes in λd trials if the probability of success in each is $1/d$. If we assume that the trials are independent, then this is a binomial random variable with parameters $n = \lambda d$ and $p = 1/d$.

Comment on parameters.

The pmf of a random variable is a function of the value of the variable. If the outcomes are numbers k, then the pmf is $f(k) :=$ the probability of outcome $\{k\}$. In the case of the binomial distribution (and other families, such as the hypergeometric and the geometric), we have an expression for the pmf that depends on parameters as well as outcome. In the binomial family, the parameters (n, p) refer to the number of trials and the probability of success on each. In the the hypergeometric family, the parameters are N, the size of the population from which a sample is drawn, s, the number of special units in the population and n, the size of the sample. In the geometric family, there is only one parameter—p, the chance of success in each trial. Each specific choice of parameter(s) determines a specific pmf from the family.

The formula for the binomial contains explicit reference to the parameters:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

We sometimes include reference to the parameters in the abbreviation:

$$P(X = k \mid n, p\,) = \binom{n}{k} p^k (1-p)^{n-k}.$$

This notation is very common in statistics. The statistician is typically interested in finding the values of the parameters of a random variable that make it into an optimal model of real data. For example, suppose a test consists in repeatedly dropping a part from a height of 1 meter until it breaks. The test is performed on 100 parts, and the number of drops till breakage is recorded as data. The statistician might assume that drops till breakage is modeled by a geometric distribution, and her task is to determine the value of the parameter p that is best supported by the data.

Solution, continued.

Using our new notation, see that the probability of having k chips in a particular cookie is

$$P(\,X = k \mid \lambda d, 1/d) = \binom{\lambda d}{k} (1/d)^k (1 - (1/d))^{\lambda d - k}.$$

We are trying determine what happens to this function as d gets larger and larger while λ remains fixed.

Note that the product of the first two factors is:

$$\binom{\lambda d}{k} (1/d)^k = \frac{\lambda d}{d} \cdot \frac{\lambda d - 1}{d} \cdots \frac{\lambda d - k + 1}{d} \cdot \frac{1}{k!}.$$

As d gets larger and larger, this number approaches

$$\frac{\lambda^k}{k!}.$$

The third factor can be factored further: $(1-(1/d))^{\lambda d-k} = (1-(1/d))^{-k} \cdot (1+(-1/d))^{d\lambda}$. As d gets larger and larger, $(1-(1/d))^{-k}$ gets closer and closer to 1. Meanwhile, as we recall from calculus:

$$\lim_{d \to \infty} \left(1 + \frac{-1}{d}\right)^{d\lambda} = e^{-\lambda}.$$

Putting all of this together, we see that:

$$\lim_{d \to \infty} P(\,X = k \mid \lambda d, 1/d\,) = \frac{\lambda^k}{k!} e^{-\lambda}$$

.