CLARK UNIVERSITY 1887

The Bernoulli process and discrete distributions
Math 217 Probability and Statistics

Prof. D. Joyce, Fall 2014

We've been looking at Bernoulli processes for a while, but didn't give them a name. Repeatedly flipping a coin is an example. Repeatedly tossing a die is another one if you only care whether a particular number comes up or not. Sampling with replacement is another.

The definition. A single trial for a Bernoulli process, called a Bernoulli trial, ends with one of two outcomes, one often called success, the other called failure. Success occurs with probability p while failure occurs with probability $1-p$. For a Bernoulli process, we'll always denote $1-p$ as q. Rather than use the outcomes heads and tails that are appropriate for flipping coins, will assume the outcomes are 0 and 1, where 1 denotes success and 0 denotes failure. A random variable X that has two outcomes, 0 and 1, where $P(X{=}1) = p$ is said to have a *Bernoulli distribution* with parameter p, Bernoulli(p).

The Bernoulli process consists of repeated independent Bernoulli trials with the same parameter p. Another way of expressing that is that Bernoulli trials form a random sample from the Bernoulli population. It's assumed that the probability p applies to every trial and that the outcomes of each trial are independent of all the rest. We'll use the random variable X_i for the outcome of the i^{th} trial in a Bernoulli process. The first n trials of a Bernoulli process $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ form a random sample from a Bernoulli distribution with parameter p.

The first serious development in the theory of probability was in the 1650s when Pascal and Fermat investigated the binomial distribution in the special case $p = \frac{1}{2}$. Pascal published the resulting theory of binomial coefficients and properties of what we now call Pascal's triangle. In the very early 1700s Jacob Bernoulli extended these results to general values of p. Bernoulli processes are named in honor of Jacob Bernoulli who studied them in detail and proved the first version of the Central Limit Theorem, something we'll look at later in the course.

Sampling with replacement. Sampling with replacement occurs when a set of N elements has a subset of M "preferred" elements, and n elements are chosen at random, but the n elements don't have to be distinct. In other words, after one element is chosen, it's put back in the set so it can be chosen again. Selecting a preferred element is success, and that happens with probability $p = M/N$, while selecting a nonpreferred element is failure, and that happens with probability $q = 1 - p = (N-M)/N$. Thus, sampling with replacement is a Bernoulli process.

Distributions associated to the Bernoulli process. You can ask various questions about a Bernoulli process, and the answers to these questions have various distributions.

- If you ask how many successes there will be among n Bernoulli trials, then the answer will have a *binomial distribution*, Binomial(n,p). Formally, it's the sum $X_1 + X_2 + \cdots + X_n$ of a Bernoulli sample of n i.i.d. random variables with a Bernoulli(p) distribution. We've looked at that before and determined that the probability that there are k successes is

$$\binom{n}{k} p^k q^{n-k}.$$

- If you ask how many trials it will be to get the first success, then the answer will have a *geometric distribution*, Geometric(p).

- If you ask how many trials there will be to get the r^{th} success, then the answer will have a *negative binomial distribution*, $\textsc{NegativeBinomial}(p, r)$.

- Given that there are M successes among N trials, if you ask how many of the first n trials are successes, then the answer will have a $\textsc{Hypergeometric}(N, M, n)$ distribution.

The geometric distribution. $\textsc{Geometric}(p)$.

In a Bernoulli process with probability p of success, the number of trials X it takes to get the first success has a geometric distribution. Let's determine what that probability is that the the first success occurs on the t^{th} trial.

The probability that the first success occurs on the first trial is, of course, p.

In order for the first success to occur on the second trial, the first has to be a failure, which occurs with probability $1-p$, followed immediately by a success, which occurs with probability p. Thus the first success occurs on the second trial is qp.

You can see where this is going. For the first success to occur on the t^{th} trial, there have to be $t-1$ failures followed by one success, and that says the probability of the first success on trial t is pq^{t-1}.

The negative binomial distribution.

$\textsc{NegativeBinomial}(p, r)$.

The negative binomial distribution generalizes the geometric distribution. Whereas the geometric distribution gives the probability that the first success occurs on the t^{th} trial, the negative binomial distribution gives the probability that the r^{th} success occurs on the t^{th} trial, where r is any positive integer. Thus, $\textsc{Geometric}(p) = \textsc{NegativeBinomial}(p, 1)$.

So what is that probability? Among the first $t-1$ trials there have to be exactly $r-1$ successes and $t-r$ failures, and that happens with probability $\binom{t-1}{r-1}p^{r-1}q^{t-r}$, and then the t^{th} trial has to be a success, and that happens with probability p. Therefore, the probability that the r^{th} success occurs on the t^{th} trial is

$$\binom{t-1}{r-1}p^r q^{t-r}.$$

Note: there are other parameters besides the p and r used here and in our text. It's not uncommon for a family of distributions to be parameterized in different ways. By the way, the "negative" in the name of this family comes from the -1's in the binomial coefficients. Perhaps a better name would have been depressed binomial distribution or something like that.

The hypergeometric distribution.

$\textsc{Hypergeometric}(N, M, n)$.

The hypergeometric distribution gives you a certain conditional probability that occurs in a Bernoulli process. Given that there are M successes among N trials, it tells you the probability that there are k successes among the first n trials. Here, k can be any nonnegative integer.

To compute this probability, take the sample space to the the $\binom{N}{M}$ ways that you can arrange M successes among the N trials. Note that they each have the same probability so we're dealing with uniform discrete probability here. Our answer won't involve p. Among those $\binom{N}{M}$ arrangements, how many of them have k successes among the first n trials? There are $\binom{n}{k}$ ways to place the k successes among the the first n trials, but then we have $\binom{N-n}{M-k}$ ways to place the remaining $M-k$ successes among the remaining $N-m$ trials. Therefore, given that there are M successes among N trials, the probability that there are k successes among the first n trials is

$$\frac{\binom{n}{k}\binom{N-n}{M-k}}{\binom{N}{M}}.$$

Sampling without replacement. Note that this is the same probability that that we got from sampling without replacement. When you have N balls, M of them preferred, and select n of them, that formula gives you the probability that you've selected k preferred ones.

Math 217 Home Page at `http://math.clarku.edu/~djoyce/ma217/`