Some examples of Probability Modeling

1 Coin tossing

A coin is tossed N times. How many heads are observed?

In order to answer this question, we need to make some assumptions/simplifications. For example, do we believe that the outcomes of the individual tosses influence each other? Do we assume that each toss of the coin has the same chances of heads? If we answer yes to each of these questions, and we let H be the number of heads and let p be the probability of a head on an individual toss, then we arrive at the familiar binomial distribution:

$$\Pr(H=j) = \binom{N}{j} p^j (1-p)^{N-j}$$

where j is any integer from 0 to N, and that $\Pr(H=j)=0$ otherwise. Note that we are forced to this conclusion by our assumptions of independence and of an unchanging probability.

On the other hand, we may believe that the all the tosses will be the same as the first one. In this case we have $\Pr(H=0)=1-p$, $\Pr(H=N)=p$, and $\Pr(H=j)=0$ otherwise. Again, this is forced by our assumptions.

2 The Memoryless Property

Suppose we wish to model the lifetime of a compact florescent lightbulb. A common engineering assumption about electrical devices is sometimes summarized as "new is a good as used". What does this mean probabalistically?

Let T be the lifetime. It seems reasonable to assume that T may assume any non-negative value. What is more, it seems to reasonable to assume that $\Pr(T>x)$ is a continuous function for $x \geq 0$, so that $\Pr(T>0)=1$. New is a good as used can be interpreted in terms of conditional probability:

$$\Pr(T>y) = \Pr(T>x+y|T>x)$$

for any $x \geq 0$ and $y \geq 0$. It is clear that the left hand side of the equation represents "new" and the right hand side represents "used". We will now see that "new is a good as used" determines the distribution of T up to a scale constant.

Let $g(x)=\Pr(T>x)$ for any non-negative x. From the definition of conditional probability we see that if $g(x)>0$ then

$$g(y) = \frac{g(x+y)}{g(x)}.$$

Multiplying through by $g(x)$ we arrive at the crucial relation:

$$g(x)g(y) = g(x+y).$$

In particular, $g(x)^2 = g(2x)$, and by induction, for any positive integer N, $g(Nx)=g(x)^N$. Since g is continuous on the positive integers, and $g(0)=1$, there is some $x>0$ so that $g(x)>0$. Hence, $g(x)>0$ for all positive x. What is more, since $g(x)\to 0$ as $x\to\infty$, $g(1)\in(0,1)$. Hence $g(1)=e^{-\lambda}$ for some $\lambda>0$. We now can show that $g(x)=g(1)^x=e^{-\lambda x}$ for all $x\geq 0$. We already know this if x is a positive integer. Now suppose $x=1/N$ where N

is a positive integer. Then since $g(1) = g(1/N)^N$, we have $g(1/N) = g(1)^{1/N}$. Hence if p and q are positive integers:

$$g(p/q) = g(1/q)^p = g(1)^{p/q} = e^{-\lambda p/q}.$$

Finally since g continuous on the non-negative real numbers and every non-negative real number is the limit of a sequence of positive rational numbers, $g(x) = e^{-\lambda x}$ for every non-negative real number x.

2.1 More lightbulbs

Suppose that we now keep track of the number of lightbulbs we install, assuming that each has an exponential lifetime with the same parameter λ. Let T_k be lifetime of the k^{th} bulb, and let N_t be the number of bulbs installed by time t. It is clear that $\{N_t \leq n\} = \{T_1 + \cdots + T_{n+1} > t\}$ for any non-negative integer n. On the other hand, it is easy to prove by induction that

$$\Pr(T_1 + \ldots + T_{n+1} > t) = \frac{1}{n!}\int_t^\infty (\lambda x)^n e^{-\lambda x} \lambda dx$$

and by integration by parts that

$$\Pr(N_t \leq n) = \sum_{k=0}^{n} \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Hence N_t must have a Poisson distribution!

3 Random Points

Consider the number of claims submitted to an insurance company in a fixed period of time. Since the callers don't know one another, it is reasonable to assume that the number of claims in disjoint time periods are independent of one another, and that the distribution of claims in any time interval depends only on the length of the interval. Finally, it would be reasonable that the claims cannot come simultaneously. We can model this by letting $C_{(a,a+h]}$ be the number of claims in $(a, a+h]$ and assuming that $\Pr(C_{(a,a+h]} = 1)/h \to \lambda 0$ as $h \to 0^+$ while $\Pr(C_{(a,a+h]} > 1)/h \to 0$ as $h \to 0^+$. This forces a unique family of probability distributions for $C_{(0,t]}$ for any positive t. Here is why.

For any positive integer N, and each integer k from 0 to N, let $t_k = kt/N$. Suppose now that we want to find $\Pr(C_{(0,t]} = a)$ for any non-negative integer a. Suppose that $N > a$, and consider the N random variables $C_{(t_{k-1},t_k]}$ for $k = 1, \ldots, N$. Then

$$\Pr(C_{(0,t]} = a) \geq \binom{N}{a} \Pr(C_{(0,t/N]} = 1)^a \left(1 - \Pr(C_{(0,t/N]} = 1\right)^{N-a}$$

Let $q(t) = \Pr(C_{(0,t]} = 1)$ and $p(t) = \Pr(C_{(0,t]} \neq 1)$.

$$\begin{aligned}\binom{N}{a} q(t/N)^a p(t/N)^{N-a} &= \frac{t^a}{a!}\frac{N!}{N^a(N-a)!}\left(\frac{q(t/N)}{t/N}\right)^a \left(1 - \frac{\lambda t}{N}\frac{q(t/n)}{\lambda t/N}\right)^{N-a} \\ &\to \frac{(\lambda t)^a}{a!}\exp(-\lambda t)\end{aligned}$$

as $N \to \infty$. These are the Poisson probabilities. We must have equality rather than $\geq$ as one probability mass function can never be greater than another without equality, as both must sum to 1. In other words, we must have

$$\Pr(C_{(0,t]} = a) = \frac{(\lambda t)^a}{a!} \exp(-\lambda t)$$

for all non-negative integers a and all $t > 0$.

4 Radioactive Decay

Suppose now that we have N plutonium atoms. Assume that they decay independently of one another, and that each has a lifetime that is distributed exponentially, or, if you like, that they follow "new is as good as used". How long do we have to wait until there are $N-1$ of them? Let T_n be the lifetime of the n^{th} one of them, and let S_1 be the time that the first one decays. Then

$$\Pr(S_1 > t) = \Pr(\min_{1 \leq n \leq N} T_n > t) = e^{-N\lambda t}.$$

In other words, the time to having $N-1$ is again exponentially distributed. What is more, the memoryless property says that the additional time needed to reduce the number to $N-2$, call this S_2, is again exponentially distributed and is independent of S_1. What is more, $\Pr(S_2 > t) = e^{-(N-1)\lambda t}$. If N_t is the number of atoms remaining at time t we have $\{N_t \leq n\} = \{S_1 + \cdots + S_{n+1} > t\}$. However, unlike our lightbulb example, it is not so easy to compute $\Pr(N_t = n)$ since the S_n are not identically distributed.