Random variables

March 11, 2009

1 Random variables

A random variable is a function that maps the sample space $\Omega$ to real numbers—namely, associates every outcome in $\Omega$ to some real number. When we assign probabilities to the sample space, we therefore automatically assign probabilities over random variables. The following example illustrates it.

Example 1. Suppose we choose three numbers without replacement from \{1, \ldots, 20\}, each choice of three numbers being just as likely as any other. Examples of random variables are $M$, the maximum of the three numbers; the sum of the three numbers; minimum $m$ of the three numbers; the median of the numbers, and so on. As we saw in class,

$$P(X = i) = \frac{(i-1)}{20} \cdot \frac{19}{20},$$

where $3 \leq i \leq 20$.

HW 1 This is the counting problem for this week. In the above example, what is the probability $m = i$, where $m$ is the minimum of the three numbers chosen?

We will consider several examples of random variables, and when it is appropriate to use them for modeling problems.

1.1 Bernoulli variables

This is perhaps the simplest of them all, and we use it to model the success or failure of an experiment. Therefore, this variable takes on only two values, 0 or 1. A Bernoulli $p$ variable assigns probability $p$ to success. Typically, we used some version of $X$ for Bernoulli variables.

Suppose we perform $n$ Bernoulli trials. An important case is when each trial is independent of the others. We looked at the coupon collector problem, and in particular, we considered the union bound. This example will be solved in several other ways in the future, and it will help to compare different solutions.

Example 2. Coupon collector problem: Suppose cereal boxes have one of $m$ different varieties coupons hidden in them. Each cereal box bought is equally likely to contain one of the $m$ varieties independently of everything else. There is a price for collecting all $m$ varieties of coupons. Suppose we buy $N$ boxes—what is the probability we have all $m$ varieties of coupons?

Let $X_j^i$ be the Bernoulli random variable that indicates if the $j$'th box has coupon variety $i$. From the problem setup, for all $1 \leq i \leq m$ and $1 \leq j \leq N$,

$$P(X_j^i = 1) = \frac{1}{m},$$

If $X_j^1 = 1$, it automatically follows that the $j$'th box has coupon variety 1, hence $X_j^i = 0$ for all $2 \leq i \leq m$. In other words, for any $j$ and for $2 \leq i \leq m$,

$$P(X_j^i = 1 | X_j^1 = 1) = 0 \neq P(X_j^i = 1),$$

namely for any $j$, the random variables $X_j^1, X_j^2, \ldots, X_j^m$ are not independent.
However, for any \( i \), \( X_i^1, X_i^2, \ldots, X_i^N \) are independent Bernoulli trials. Let
\[
B^{(i)} = \{X_i^1 = 0 \text{ and } X_i^2 = 0, \ldots, X_i^N = 0\},
\]  
be the set of all ways we can purchase \( N \) cereal-boxes and never get coupon \( i \). Can we compute the probability of \( B^{(i)} \)? Indeed we can. From (1), it follows that
\[
P(B^{(i)}) = P(X_i^1 = 0, X_i^2 = 0, \ldots, X_i^N = 0)
\]
\[
= P(X_i^1 = 0)P(X_i^2 = 0 | X_i^1 = 0) \cdots P(X_i^N = 0 | X_i^1 = 0, X_i^2 = 0, \ldots, X_i^{N-1} = 0)
\]
\[
= P(X_i^1 = 0)P(X_i^2 = 0) \cdots P(X_i^N = 0)
\]
\[
= \left(1 - \frac{1}{m}\right)^N,
\]
where the last but one equality follows since \( X_i^1, X_i^2, \ldots, X_i^N \) are independent Bernoulli trials (each box is independent of the other boxes).

We want every coupon, so we do not like any of the \( B^{(i)} \). In particular,
\[
B = B^{(1)} \cup B^{(2)} \cup B^{(3)} \ldots \cup B^{(m)}
\]
denotes the set of all purchases of \( N \) cereal-boxes where we miss either coupon 1 or coupon 2 or coupon 3 or... coupon \( m \), namely all purchases that missed a coupon.

To bound the probability of the bad event, \( B \), we notice that it is possible for a purchase to miss both coupons 1 and 2—therefore \( B^{(1)} \) and \( B^{(2)} \) are not disjoint. In general \( B^{(i)} \) are not disjoint. We only know how to compute probabilities of disjoint events, so we upper bound the probability of \( B \) instead of computing it. The following upper bound is called the union bound, and it is terribly useful.
\[
P(B) \leq P(B^{(1)}) + P(B^{(2)}) + \ldots + P(B^{(m)}) = m \left(1 - \frac{1}{m}\right)^N,
\]
where the equality is by substituting for \( P(B^{(i)}) \). As we saw in class, \( 1 - x \leq e^{-x} \), therefore,
\[
P(B) \leq me^{-N/m}.
\]
If we pick a number \( c > 0 \), and \( N = m \log m + \alpha m \), then
\[
P(B) \leq e^{-\alpha}.
\]
By picking \( \alpha \) large enough, we essentially can make the probability of the bad event \( B \) as small as we want.

Let
\[
G = B^c
\]
denote the set of all purchases of \( N \) cereal-boxes that contain at least one instance of every variety of coupon.

Then the probability we collect all varieties of coupons in \( N = m \log m + \alpha m \) boxes,
\[
P(G) = P(B^c) \geq 1 - e^{-\alpha},
\]
can be made as close to 1 as we want.

\[\Box\]

**HW 2** Suppose we are interested only in coupon of variety 1. What is the probability we will obtain at least one coupon of variety 1 among the \( N \) boxes purchased?

### 1.2 Binomial variables

Binomial variables model the number of successes in independent Bernoulli trials. Given \( n \) Bernoulli trials, \( X_1, \ldots, X_n \), each independently successful with probability \( p \), the number of successes \( Y \) is a Binomial
Consider one sequence of Resolve the answers of parts 2 and 3 in this problem. Namely, explain why it is ok if they are different.

Show that the probability of all sequences with at least $p$ ones is $\frac{(n)}{k}P^k(1-p)^{n-k}$.

For $Y$ to be $k$, we have to succeed in exactly $k$ trials. The following events correspond to the probabilities of $Y = 0, 1, \ldots, n$.

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$\ldots$</th>
<th>$X_n$</th>
<th>$P(X_1, \ldots, X_n)$</th>
<th>$#(X_1, \ldots, X_n)$ for $Y$</th>
<th>$P(Y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$\ldots$</td>
<td>0</td>
<td>$(1-p)^n$</td>
<td>$(\binom{n}{0}) = 1$</td>
<td>$(1-p)^n$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>$\ldots$</td>
<td>0</td>
<td>$p(1-p)^{n-1}$</td>
<td>$n$</td>
<td>$np(1-p)^{n-1}$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>$\ldots$</td>
<td>0</td>
<td>$p(1-p)^{n-1}$</td>
<td>$n$</td>
<td>$np(1-p)^{n-1}$</td>
</tr>
<tr>
<td></td>
<td>$\ldots$</td>
<td>$\vdots$</td>
<td>$\ldots$</td>
<td>$\ldots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>$\ldots$</td>
<td>1</td>
<td>$p(1-p)^{n-1}$</td>
<td>$n$</td>
<td>$np(1-p)^{n-1}$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>$\ldots$</td>
<td>0</td>
<td>$p^2(1-p)^{n-2}$</td>
<td>$(\binom{n}{2})$</td>
<td>$(\binom{n}{2})p^2(1-p)^{n-2}$</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>$\ldots$</td>
<td>1</td>
<td>$p^2(1-p)^{n-2}$</td>
<td>$(\binom{n}{2})$</td>
<td>$(\binom{n}{2})p^2(1-p)^{n-2}$</td>
</tr>
<tr>
<td></td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>$\ldots$</td>
<td>1</td>
<td>$p^2(1-p)^{n-2}$</td>
<td>$(\binom{n}{2})$</td>
<td>$(\binom{n}{2})p^2(1-p)^{n-2}$</td>
</tr>
<tr>
<td>$k$</td>
<td>1</td>
<td>1</td>
<td>$\ldots$</td>
<td>0</td>
<td>$p^k(1-p)^{n-k}$</td>
<td>$(\binom{n}{k})$</td>
<td>$(\binom{n}{k})p^k(1-p)^{n-k}$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$k$</td>
<td>1</td>
<td>0</td>
<td>$\ldots$</td>
<td>1</td>
<td>$p^k(1-p)^{n-k}$</td>
<td>$(\binom{n}{k})$</td>
<td>$(\binom{n}{k})p^k(1-p)^{n-k}$</td>
</tr>
<tr>
<td>$n$</td>
<td>1</td>
<td>1</td>
<td>$\ldots$</td>
<td>1</td>
<td>$p^n$</td>
<td>$(\binom{n}{n}) = 1$</td>
<td>$p^n$</td>
</tr>
</tbody>
</table>

There are $\binom{n}{k}$ different ways (# rows that correspond to $Y = k$), we could be successful in exactly $k$ of $n$ trials. Consider one of the $\binom{n}{k}$ ways in which we are successful in exactly $k$ of $n$ trials. Say we are successful in the first $k$ and fail the other $n-k$ trials. The probability we are successful in all of the first $k$ trials and fail every of the last $n-k$ trials is $p^k(1-p)^{n-k}$.

No matter which $k$ of the $n$ trials we choose to be successful in (no matter which row we choose when $Y = k$), the probability of $k$ successes and $n-k$ failures is $p^k(1-p)^{n-k}$. Therefore $P(Y = k) = \binom{n}{k}p^k(1-p)^{n-k}$.

**HW 3** Suppose you have $n$ independent Bernoulli $p$ trials.

1. What is the probability you will pick $k$ ones? Call this $P(k)$.

2. Let $p < \frac{1}{2}$. For what value of $k$ is $P(k)$ (see previous subproblem for definition of $P(k)$) maximized? To do this part, you can use the following trick: compute the ratio $P(k+1)/P(k)$. For what values is the ratio bigger than 1? Hint: Solve the following inequality for $k$,

$$\frac{P(k + 1)}{P(k)} > 1,$$

and thus obtain the range of $k$ for which the equation above is true. For what values is the ratio less than 1? Therefore, what is the value of $k$ so that $P(k)$ is maximum? You can assume $np$ is an integer if it helps.

3. Consider one sequence of $n$ bits with $k$ ones and $n-k$ zeros. What is its probability? If $p < 1/2$, what sequence has the highest probability?

4. Resolve the answers of parts 2 and 3 in this problem. Namely, explain why it is ok if they are different.

5. Show that the probability of all sequences with at least $k$ ones is $\geq p^k$, namely,

$$\sum_{i=k}^{n} P(i) \geq p^k.$$
Hint: lower bound $\sum_{i=k}^n P(i)$ by the probability of the set $S_k$ of sequences whose first $k$ positions are 1 (other positions can be either 1 or 0). Then show $P(S_k) = p^k$. Namely, if $X_i$ is the bit at position $i$,

$$\sum_{i=k}^n P(i) \geq P(X_1 = 1, \ldots, X_k = 1, X_{k+1} = 1 \text{ or } 0, \ldots, X_n = 1 \text{ or } 0)$$

$$= P(X_1 = 1, \ldots, X_k = 1)$$

$$= p^k.$$

### 1.2.1 Binomial theorem

Note that $\sum_{k=0}^n P(Y = k) = 1$, so it follows that

$$\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = 1. \quad (2)$$

This is a special case of what is known as the Binomial Theorem. In general for all (real, complex) $x$ and $y$,

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}.$$ 

Setting $x = p$ and $y = 1 - p$, we obtain (2).

### 1.3 Poisson variables

Poisson random variables can take up any value from the set $\{0, 1, \ldots\}$, namely any non-zero number. Poisson random variables take in one parameter $\lambda$. Let $Z$ be a Poisson random variable with parameter $\lambda$.

We write $Z \sim \text{Poisson}(\lambda)$, and for $k = 0, 1, \ldots$, the probability that $Z = k$ is

$$P(Z = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$ 

Poisson random variables approximate binomial probabilities under certain conditions. Specifically a Binomial variable corresponding to the number of successes in $n$ independent Bernoulli trials each having probability $p$ each can be well approximated by a Poisson random variable $Z$ with parameter $np$ if $np \ll 1$.

**HW 4** We will prove that if $Y \sim \text{Bin}(n, p)$ and $Z \sim \text{Poisson}(np)$, then we can approximate the $P(Y = k)$ by $P(Z = k)$.

1. If $Y \sim \text{Bin}(n, p)$, then show that

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} \leq \frac{n^k}{k!} p^k (1-p)^{n-k}$$

2. Next show using $1 - x \leq e^{-x}$ that

$$\frac{n^k}{k!} p^k (1-p)^{n-k} \leq \frac{n^k}{k!} p^k e^{-np} e^{kp},$$

Therefore, show that if $0 \leq np < 1$,

$$\frac{n^k}{k!} p^k (1-p)^{n-k} \leq \frac{n^k}{k!} p^k e^{-np} \frac{1}{1 - kp} \leq \frac{n^k}{k!} p^k e^{-np} \frac{1}{1 - np},$$

thus concluding that

$$P(Y = k) \leq \frac{P(Z = k)}{1 - np}.$$ 

A lower bound can also be derived (with more work). If $np$ is .001, what is the ratio of the above upper bound on $P(Y = k)$ to $P(Z = k)$?
1.4 Geometric random variable

The geometric random variable $G$ with parameter $p$ models the instant of first success of independent Bernoulli $p$ trials. Let $X_i$ denote the $i$'th Bernoulli trial. Therefore geometric random variables can take any positive number as a value, and for all $k \geq 1$

$$P(G = k) = P(X_1 = 0, X_2 = 0, X_{k-1} = 0, X_k = 1)$$
$$= P(X_1 = 0)P(X_2 = 0|X_1 = 0)P(X_3 = 0|X_1 = 0, X_2 = 0)\ldots P(X_k = 1|X_1 = 0, X_2 = 0, \ldots, X_{k-1} = 0)$$
$$= P(X_1 = 0)P(X_2 = 0)P(X_3 = 0)\ldots P(X_k = 1)$$
$$= (1 - p)^{k-1}p.$$