Chapter 3

Things binomial

The standard coin-tossing mechanism drives much of classical probability. It generates several standard distributions, the most important of them being the Binomial. The distributions appear often in probabilistic modelling; it is worthwhile recording a few of their properties.

As a probabilist, I tend to regard any method involving probability calculations as a vast improvement over purely analytic methods. That will be my excuse for the following calculation.

<3.1> **Example.** How many ways are there to choose a subset of size k from a set of n objects, for $k = 0, 1, \dots, n$? It is traditional to write this number as $\binom{n}{k}$, read "n choose k." By convention, $\binom{n}{0} = 1$. I'll use a conditional probability argument to find $\binom{n}{k}$ for $k \geq 1$.

Consider a slightly different question. Suppose the objects are numbered $1, 2, \dots, n$. Choose a subset of size k "at random." What is the probability that it consists precisely of objects 1 to k? Calculate the result in two ways, then equate the answers.

Method I.

Interpret "at random" to mean that all $\binom{n}{k}$ possible subsets of size k are equally likely, so that $\mathbb{P}\{\text{choose 1 to k}\} = 1/\binom{n}{k}$.

Method II.

Generate the random k-set one member at a time: choose the first member at random from the n available objects; then choose the second member at random from the remaining $n-1$ objects; and so on. Is it obvious that all k-sets have equal probability of being chosen? Write F_i for the event {the ith choice is one of 1,2,...,k}. Then

$$
\begin{aligned}
\mathbb{P}\{\text{choose 1 to } k\} &= \mathbb{P}F_1F_2\dots F_k \\
&= \mathbb{P}F_1\mathbb{P}\big(F_2 \mid F_1\big)\dots\mathbb{P}\big(F_k \mid F_1F_2\dots F_{k-1}\big) \\
&= \frac{k}{n}.\frac{k-1}{n-1}.\frac{k-2}{n-2}\dots\frac{1}{n-k+1} \\
&= \frac{k!(n-k)!}{n!}.
\end{aligned}
$$

Notice how the $(n-k)!$ cancels out all except the k largest factors in $n!$. Equate the two solutions to get

$$
\binom{n}{k} = \frac{n!}{k!(n-k)!}
$$

□

•binomial coefficient

The symbol $\binom{n}{k}$ is called a BINOMIAL COEFFICIENT because of its connection with the binomial expansion:

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

The expansion can be generalized to fractional and negative powers by means of Taylor's theorem. For general real α define

$$\binom{\alpha}{0} = 1 \qquad \text{and} \qquad \binom{\alpha}{k} = \frac{\alpha(\alpha-1)(\alpha-2)\dots(\alpha-k+1)}{k!} \qquad \text{for } k = 1, 2, \dots$$

Then

$$(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k \qquad \text{at least for } |x| < 1.$$

The Binomial distribution arises in any situation where one is interested in the number of successes in a fixed number of independent trials (or experiments), each of which can result in either success or failure.

<3.2> **Example.** For n independent tosses of a coin that lands heads with probability p, find

(i) the distribution of X, the total number of heads

(ii) the expected value of X

Clearly X can take only values $0, 1, 2, \dots, n$. For a fixed a k in this range, break the event $\{X = k\}$ into disjoint pieces like

$$F_1 = \{\text{first k gives heads, next n-k give tails}\}$$
$$F_2 = \{\text{first (k-1) give heads, then tail, then head, then n- k-1 tails}\}$$
$$\vdots$$

The indexing on the F_i is most uninformative. (Maybe you can think of something better.) It matters only that each F_i specifies k positions for the heads and leaves the remaining $n-k$ for tails. Write H_j for {jth toss is a head}. Then

$$\begin{aligned} \mathbb{P}F_1 &= \mathbb{P}\big(H_1 H_2 \dots H_k H_{k+1}^c \dots H_n^c\big) \\ &= (\mathbb{P}H_1)(\mathbb{P}H_2)\dots(\mathbb{P}H_n^c) \qquad \text{by independence} \\ &= p^k(1-p)^{n-k}. \end{aligned}$$

A similar calculation gives $\mathbb{P}F_i = p^k(1-p)^{n-k}$ for every other i; all that changes is the order in which the p and $(1-p)$ factors appear. From the previous Example there are exactly $\binom{n}{k}$ different F_i's, because each F_i corresponds to a different choice of the k positions for the heads to occur. Adding up that many of the $p^k(1-p)^{n-k}$ probabilities, we get

$$\mathbb{P}\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{for } k = 0, 1, \dots, n.$$

A random variable that takes these values with these probabilities is said to have a "binomial distribution with parameters n and p," or Bin(n, p) distribution, for short.

For part (ii) there are hard ways and easy ways to proceed.

Hard way: By the formula in Chapter 2,

$$\mathbb{E}X = \sum k = 0^n k \binom{n}{k} p^k (1-p)^{n-k} = \ ??$$

The series is not so hard to sum, but why try?

Easy way: Use the method of indicators, as in Chapter 2. Define

$$X_i = \begin{cases} 1 & \text{if ith toss is head} \\ 0 & \text{if ith toss is tail.} \end{cases}$$

Then $X = X_1 + \dots X_n$ and $\mathbb{E}X = \mathbb{E}X_1 + \dots \mathbb{E}X_n$ by multiple applications of rule E1 for expectations. Consider X_1. From rule E4,

$$\mathbb{E}X_1 = 0\mathbb{P}\{X_1 = 0\} + 1\mathbb{P}\{X_1 = 1\} = p.$$

□ Similarly $\mathbb{E}X_i = p$ for all the other X_i. Add to get $\mathbb{E}X = np$.

•marginal

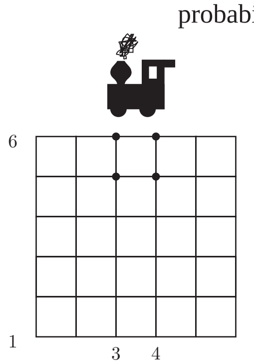
The calculation for part (ii) made no use of the independence. If each X_i has MARGINAL distribution Bin(1,p), that is, if

$$\mathbb{P}\{X_i = 1\} = p = 1 - \mathbb{P}\{X_i = 0\} \qquad \text{for each } i,$$

then $\mathbb{E}(X_1 + \ldots X_n) = np$, regardless of possible dependence between the tosses.

<3.3> **Example.** An unwary visitor to the Big City is standing at the corner of 1st Street and 1st Avenue. He wishes to reach the railroad station, which actually occupies the block on 6th Street from 3rd to 4th Avenue. (The Street numbers increase as one moves north; the Avenue numbers increase as one moves east.) He is unaware that he is certain to be mugged as soon as he steps onto 6th Street or 6th Avenue.

Being unsure of the exact location of the railroad station, the visitor lets himself be guided by the tosses of a fair coin: at each intersection he goes east with probability 1/2 and north with probability 1/2. What is the probability that he is mugged outside the railroad station?



To get mugged at (3,6) or (4,6) the visitor must proceed north from either the intersection (3,5) or the intersection (4,5)—we may assume that if he gets mugged at (2,6) and then moves east, he won't get mugged again at (3,6), which would be an obvious waste of valuable mugging time for no return. The two possibilities correspond to disjoint events.

$$\begin{aligned}
&\mathbb{P}\{\text{mugged at railroad}\} \\
&\quad = \mathbb{P}\{\text{reach (3,5), move north}\} + \mathbb{P}\{\text{reach (4,5), move north}\} \\
&\quad = {}^1\!/_2\mathbb{P}\{\text{reach (3,5)}\} + {}^1\!/_2\mathbb{P}\{\text{reach (4,5)}\} \\
&\quad = {}^1\!/_2\mathbb{P}\{\text{move east twice during first 6 blocks}\} \\
&\qquad\quad + {}^1\!/_2\mathbb{P}\{\text{move east 3 times during first 7 blocks}\}.
\end{aligned}$$

A better way to describe the last event might be "move east 3 times and north 4 times, in some order, during the choices governed by the first 7 tosses of the coin." The Bin(7,1/2) lurks behind the calculation. The other calculation involves the Bin(6,1/2).

$$\mathbb{P}\{\text{mugged at railroad}\} = \frac{1}{2}\binom{6}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^4 + \frac{1}{2}\binom{7}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^4 = \frac{65}{256}.$$

□

<3.4> **Example.** Suppose a multiple-choice exam consists of a string of unrelated questions, each having three possible answers. Suppose there are two types of candidate who will take the exam: guessers, who make a blind stab on each question, and skilled candidates, who can always eliminate one obviously false alternative, but who then choose at random between the two remaining alternatives. Suppose 70% of the candidates who take the exam are skilled and the other 30% are guessers. A particular candidate has gotten 4 of the first 6 question correct. What is the probability that he will also get the 7th question correct?

Interpret the assumptions to mean that a guesser answers questions independently, with probability 1/3 of being correct, and that a skilled candidate also answers independently, but with probability 1/2 of being correct. Let X denote the number of questions answered correctly from the first six. Then

(i) for a guesser, X has (conditional) distribution Bin(6,1/3)

(ii) for a skilled candidate, X has (conditional) distribution Bin (6,1/2).

Let G denote the event {the candidate is a guesser} and S denote the event {the candidate is skilled}. We are to assume that

$$\mathbb{P}G = 0.3 \text{ and } \mathbb{P}S = 0.7.$$

The question asks for $\mathbb{P}\{\text{next correct} \mid X = 4\}$.

Split according to the type of candidate, then condition.

$$
\begin{aligned}
&\mathbb{P}\{\text{next correct} \mid X = 4\} \\
&\quad = \mathbb{P}\{\text{next correct, } S \mid X = 4\} + \mathbb{P}\{\text{next correct, } G \mid X = 4\} \\
&\quad = \mathbb{P}(S \mid X = 4)\mathbb{P}\{\text{next correct} \mid X = 4, S\} + \mathbb{P}(G \mid X = 4)\mathbb{P}\{\text{next correct} \mid X = 4, G\}.
\end{aligned}
$$

If we know the type of candidate, the $\{X = 4\}$ information becomes irrelevant, reducing the last expression to

$$\tfrac{1}{2}\mathbb{P}(S \mid X = 4) + \tfrac{1}{3}\mathbb{P}(G \mid X = 4).$$

Notice how the success probabilities are weighted by probabilities that summarize our current knowledge about whether the candidate is skilled or guessing. If the roles of $\{X = 4\}$ and type of candidate were reversed we could use the conditional distributions for X to calculate conditional probabilities:

$$
\begin{aligned}
\mathbb{P}(X = 4 \mid S) &= \binom{6}{4}(1/2)^4(1/2)^2 2 = \binom{6}{4}\,1/64 \\
\mathbb{P}(X = 4 \mid G) &= \binom{6}{4}(1/3)^4(2/3)^2 = \binom{6}{4}\,4/729.
\end{aligned}
$$

I have been lazy with the binomial coefficients because they will later cancel out.

Apply the usual splitting/conditioning argument.

$$
\begin{aligned}
\mathbb{P}(S \mid X = 4) &= \frac{\mathbb{P}S\{X = 4\}}{\mathbb{P}\{X = 4\}} \\
&= \frac{\mathbb{P}(X = 4 \mid S)\mathbb{P}S}{\mathbb{P}(X = 4 \mid S)\mathbb{P}S + \mathbb{P}(X = 4 \mid G)\mathbb{P}G} \\
&= \frac{\binom{6}{4}\,1/64(.7)}{\binom{6}{4}\,1/64(.7) + \binom{6}{4}\,4/729(.3)} \\
&\approx .869.
\end{aligned}
$$

There is no need to repeat the calculation for the other conditional probability, because

$$\mathbb{P}(G \mid X = 4) = 1 - \mathbb{P}(S \mid X = 4) \approx .131.$$

Thus, given the 4 out of 6 correct answers, the candidate has conditional probability of approximately

$$\tfrac{1}{2}(.869) + \tfrac{1}{3}(.131) \approx .478$$

□ of answering the next question correctly.

Some authors prefer to summarize the calculations by means of the *odds ratios*:

$$\frac{\mathbb{P}(S \mid X = 4)}{\mathbb{P}(G \mid X = 4)} = \frac{\mathbb{P}S}{\mathbb{P}G} \cdot \frac{\mathbb{P}(X = 4 \mid S)}{\mathbb{P}(X = 4 \mid G)}.$$

The initial ratio of $\mathbb{P}S/\mathbb{P}G$ is multiplied by a factor that reflects the relative support of the data for the two competing explanations "skilled" and "guessing". The conditioning calculation for $\mathbb{P}(S \mid X = 4)$ is an instance of BAYES'S FORMULA. The whole Example is an instance of BAYESIAN INFERENCE.

•Bayes's formula

<3.5> **Example.** Members of the large governing body of a small country are given special banking privileges. Unfortunately, some members appear to be abusing the privilege by writing bad checks. The royal treasurer declares the abuse to be a minor aberration, restricted to fewer than 5% of the members. An investigative reporter manages to expose the bank records of 20 members, showing that 4 of them have been guilty. How credible is the treasurer's assertion?

Suppose a fraction p of the members are guilty. If the sample size 20 is small relative to the population of members, and if the reporter was getting a representative sample, the number of guilty in the sample should be distributed Bin$(20, p)$. You should be able to think of many ways in which these assumptions could be violated, but I'll calculate as if the simple Binomial model were correct.

Write X for the number of guilty in the sample, and add a subscript p to the probabilities to show that they refer to the Bin(20,p) distribution. Before the sample is taken we could assert

$$\begin{aligned}&\mathbb{P}_p\{X \geq 4\}\\ &\quad= \binom{20}{4}p^4(1-p)^{16} + \binom{20}{5}p^5(1-p)^{14} + \ldots + \binom{20}{4}p^{20}(1-p)^0\\ &\quad= 1 - \left(\binom{20}{0}p^0(1-p)^{20} + \binom{20}{1}p^1(1-p)^{19} + \binom{20}{2}p^2(1-p)^{18} + \binom{20}{3}p^3(1-p)^{17}\right).\end{aligned}$$

The second form makes it easier to calculate by hand when $p = .05$:

$$\mathbb{P}_{.05}\{X \geq 4\} \approx .02.$$

For values of p less than 0.05 the probability is even smaller.

●hypothesis testing

After the sample is taken we are faced with a choice: either the treasurer is right, and we have just witnessed something very unusual; or maybe we should disbelieve the 5% upper bound. This dichotomy illustrates the statistical procedure called HYPOTHESIS TESTING. One chooses an event that should be rare under one model, but more likely under an alternative model, as a guide to a simple *believe model/don't believe model* response to an experiment. For the present example the event $\{X \geq 4\}$ would have been much more likely under alternative explanations involving larger proprtions of bad-check writers amongst the members.

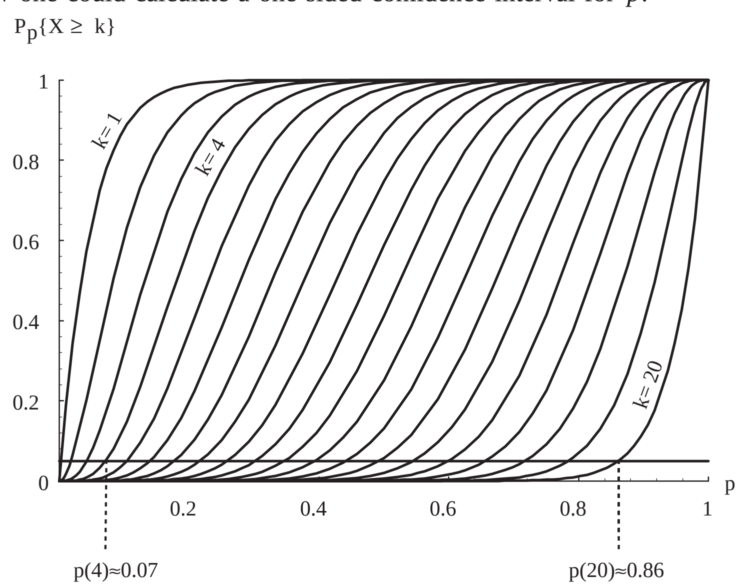
You could safely skip the remainder of this Example. It discusses a concept from theoretical statistics as an excuse to make more calculations with Binomial distributions.

✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠ ✠

Sometimes a simple yes/no response is inadequate. Given the nature of X, one would like a plausible range of values for p. More specifically, given $X = 4$, what would be a reasonable lower bound for possible p values?

●confidence interval

Many statisticians would quote a CONFIDENCE INTERVAL for p in response to the last question. The interpretation is subtle; the interval does not carry the meaning that one might assume. (Some statisticians of the Bayesian persuasion have been unkind enough to point out similarities between confidence intervals and confidence tricks.) With this encouraging introduction, let me explain how one could calculate a one-sided confidence interval for p.



Remember that $\mathbb{P}_p\{\ldots\}$ refers to calculations under which X has a Bin(20,p) distribution. For each k the probability $\mathbb{P}_p\{X \geq k\}$ is increasing as a function of p. If $1 \leq k \leq 20$, it increases smoothly from 0 to 1 as p increases from 0 to 1. With some small effort one can solve for the values

$$0 < p(1) < p(2) < \ldots < p(20) < 1$$

for which

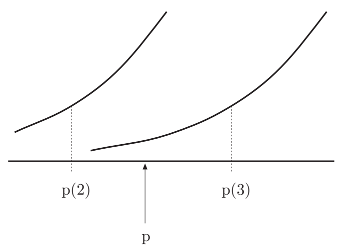
$$\mathbb{P}_{p(k)}\{X \geq k\} = 0.05.$$

Define $p(0) = 0$, so that $p(X)$ is well defined for all possible values of X.

Let C denote the random interval $[p(X), 1]$. I assert that C has the property

<3.6>

$$\mathbb{P}_p\{C \text{ contains } p\} \geq .95 \qquad \text{for every } p.$$



To see why the inequality holds, consider a typical p. Suppose, for example, that $p(2) \leq p < p(3)$. The random interval $C = [p(X), 1]$ fails to contain p if $p < p(X)$. That happens if X takes a value k for which $p < p(k)$, which, in the present case, holds for $k = 3, 4, \ldots, 20$. Similarly, the interval C contains p if X takes a value k for which $p(k) \leq p$; it contains p if X takes values 0,1, or 2. Thus

$$\begin{aligned}\mathbb{P}_p\{C \text{ does not contain } p\} &= \mathbb{P}_p\{X \geq 3\} \\ &\leq \mathbb{P}_{p(3)}\{X \geq 3\} \qquad \text{because } \mathbb{P}_p\{X \geq 3\} \text{ increases with } p \\ &= 0.05 \qquad \text{by definition of } p(3).\end{aligned}$$

Subtract both sides of the inequality from 1 to get <3.6>, at least for p between $p(2)$ and $p(3)$. A similar argument establishes <3.6> for the other ranges of p.

Now for the subtle part. If the reporter observes $X = 4$ he would calculate $[p(4), 1]$ as the one-sided confidence interval, perhaps announcing that he is 95% confident that the unknown p lies in the range 0.07 to 1. (The value of $p(4)$ is approximately 0.07.) What does that mean? It does not mean that p has probability 0.95 of lying in the range $[0.07, 1]$. A statistician who accepts the frequency interpretation might explain:

> "There is a fixed value of p that we don't know. Maybe it is greater than 0.07, and maybe it's not. Who knows? But if you keep taking samples of size 20 and calculating the intervals $[p(X), 1]$, in about 95% of the samples you will actually cover the unknown p."

□ Now you know.