Chapter 3

Things binomial

3.1 Overview

The standard coin-tossing mechanism drives much of classical probability. It generates several standard distributions, the most important of them being the Binomial. The name comes from the ***binomial coefficient***, $\binom{n}{k}$, which is defined as the number of subsets of size k for a set of size n. (Read the symbol as "n choose k".) Clearly, $\binom{n}{0} = 1 = \binom{n}{n}$: there is only one empty subset and only one subset containing everything.

Here is a neat probabilistic way to determine $\binom{n}{k}$, for integers $1 \le k \le n$. Suppose k balls are sampled at random, without replacement, from an urn containing k red balls and $n-k$ black balls. Each of the $\binom{n}{k}$ different subsets of size k has probability $1/\binom{n}{k}$ of being selected. In particular, the event

$$A = \{\text{the sample consists of the red balls}\}$$

has probabilty $1/\binom{n}{k}$. We can also calculate this probability using a conditioning argument. Given that the first i balls are red, the probability that the $(i+1)$st is red is $(k-i)/(n-i)$. Thus

$$\mathbb{P}A = \frac{k}{n}.\frac{k-1}{n-1}.\frac{k-2}{n-2}\dots\frac{1}{n-k+1} = \frac{k!(n-k)!}{n!}.$$

Equating the two values for $\mathbb{P}A$ we get

$$\binom{n}{k} = \frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}$$

The formula also holds for $k = 0$ if we interpret $0!$ as 1.

Remark. 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

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

Then

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

Definition. (Binomial distribution) A random variable X is said to have a $\text{Bin}(n,p)$ distribution, for a parameter p in the range $0 \le p \le 1$, if it can take values $0, 1, \ldots, n-1, n$ with probabilities

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

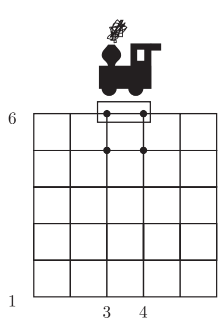
Compare with the binomial expansion,

$$1 = (p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \qquad \text{where } q = 1-p.$$

Example <3.1> For n independent tosses of a coin that lands heads with probability p, show that the total number of heads has a $\text{Bin}(n,p)$ distribution, with expected value np.

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.

Example <3.2> 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, or north, with probability 1/2. What is the probability that he is mugged outside the railroad station?

The following problem is an example of ***Bayesian inference***, based on the probabilistic result known as ***Bayes's rule***. You need not memorize the rule, because it is just an application of the conditioning method you already know.

Example <3.3> Suppose a multiple-choice exam consists of a string of unrelated questions, each having three possible answers. Suppose also that 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. Finally, 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?

As a method of solving statistical problems, Bayesian inference is advocated devoutly by some Statisticians, and derided by others. There is no disagreement regarding the validity of Bayes's rule; it is the assignment of prior probabilities—such as the $\mathbb{P}S$ and $\mathbb{P}G$ of the previous Example—that is controversial in a general setting.

The Bayesian message comes through more strongly in the next Example.

Example <3.4> Suppose we have three coins, which land heads with probabilities p_1, p_2, and p_3. Choose a coin according to the ***prior distribution*** $\theta_i = \mathbb{P}\{\text{ choose coin } i\ \}$, for $i = 1, 2, 3$, then toss that coin n times. Find the posterior probabilities $\mathbb{P}\{\text{ chose coin } i \mid k \text{ heads with } n \text{ tosses }\}$, for $k = 0, 1, \ldots, n$.

We will meet the Binomial again.

3.2 The examples

<3.1> **Example.** *For n independent tosses of a coin that lands heads with probability p, show that the total number of heads has a Bin(n, p) distribution, with expected value np.*

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

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

Here i runs from 1 to $\binom{n}{k}$, because each F_i corresponds to a different choice of the k positions for the heads to occur.

Remark. The indexing on the F_i is most uninformative; it gives no indication of the corresponding pattern of heads and tails. Maybe you can think of something better.

Write H_j for the event {jth toss is a head}. Then

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

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. Thus

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

which is the asserted Binomial distribution.

It is possible to calculate $\mathbb{E}X$ by the summation formula

$$
\begin{aligned}
\mathbb{E}X &= \sum\nolimits_{k=0}^{n} \mathbb{E}(X \mid X = k)\mathbb{P}\{X = k\} \\
&= \sum\nolimits_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum\nolimits_{k=1}^{n} \frac{n(n-1)!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= np \sum\nolimits_{k-1=0}^{n-1} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \qquad \text{cf. binomial expansion of } (p + (1-p))^{n-1}.
\end{aligned}
$$

The manipulations of the sums was only slightly tedious, but why endure even a little tedium when the method of indicators is so much simpler? Define

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

Then $X = X_1 + \dots X_n$, which gives $\mathbb{E}X = \mathbb{E}X_1 + \dots \mathbb{E}X_n = n\mathbb{E}X_1$. Calculate.

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

Thus $\mathbb{E}X = np$.

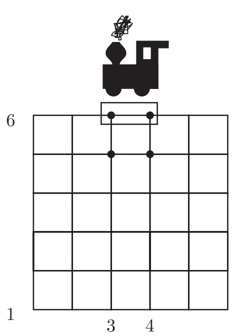
> **Remark.** The calculation of the expected value made no use of the independence. If each X_i has ***marginal*** distribution $\text{Ber}(p)$, that is, if
>
> $$\mathbb{P}\{X_i = 1\} = p = 1 - \mathbb{P}\{X_i = 0\} \qquad \text{for each } i,$$
>
> then $\mathbb{E}(X_1 + \dots X_n) = np$, regardless of possible dependence between the tosses. The expectation of a sum is the sum of the expectations, no matter how dependent the summands might be.

The symbol Ber stands for "Bernoulli".

□

<3.2> **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 $1/2$*, or 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.

$$\begin{aligned}
&\mathbb{P}\{\text{mugged at railroad}\}\\
&\quad = \mathbb{P}\{\text{reach (3,5), move north}\} + \mathbb{P}\{\text{reach (4,5), move north}\}\\
&\quad = \tfrac{1}{2}\mathbb{P}\{\text{reach (3,5)}\} + \tfrac{1}{2}\mathbb{P}\{\text{reach (4,5)}\}\\
&\quad = \tfrac{1}{2}\mathbb{P}\{\text{move east twice during first 6 blocks}\}\\
&\qquad\quad + \tfrac{1}{2}\mathbb{P}\{\text{move east 3 times during first 7 blocks}\}.
\end{aligned}$$

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).

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

> **Remark.** Notice that the events {reach (3,5)} and {reach (4,5)} are not disjoint. We need to include the part about moving north to get a clean break.

□

<3.3> **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. Let C denote the event {question 7 answered correctly}, G denote the event {the candidate is a guesser}, and S denote the event {the candidate is skilled}. Then

(i) conditional on being a guesser, X has a Bin(6, 1/3) distribution (sometimes abbreviated to $X \mid G \sim \text{Bin}(6, 1/3)$)

(ii) conditional on being a skilled candidate, X has a Bin(6, 1/2) distribution (sometimes abbreviated to $X \mid S \sim \text{Bin}(6, 1/2)$).

(iii) $\mathbb{P}G = 0.3$ and $\mathbb{P}S = 0.7$.

The question asks for $\mathbb{P}(C \mid X = 4)$.

Split according to the type of candidate, then condition.

$$\begin{aligned}\mathbb{P}(C \mid X = 4) &= \mathbb{P}\{CS \mid X = 4\} + \mathbb{P}\{CG \mid X = 4\}\\ &= \mathbb{P}(S \mid X = 4)\mathbb{P}(C \mid X = 4,\, S)\\ &\quad + \mathbb{P}(G \mid X = 4)\mathbb{P}(C \mid X = 4,\, G).\end{aligned}$$

If we know the type of candidate, the $\{X = 4\}$ information becomes irrelevant. The last expression simplifies 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:

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

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.

Remark. 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 odds ratio, $\mathbb{P}S/\mathbb{P}G$, is multiplied by a factor that reflects the relative support of the data for the two competing explanations "skilled" and "guessing".

□

<3.4> **Example.** *Suppose we have three coins, which land heads with probabilities p_1, p_2, and p_3. Choose a coin according to the* ***prior distribution*** $\theta_i = \mathbb{P}\{\text{choose coin } i\}$, *for $i = 1, 2, 3$, then toss that coin n times. Find the posterior probabilities*

$$\mathbb{P}\{\text{chose coin } i \mid \ k \text{ heads with } n \text{ tosses }\} \qquad \text{for } k = 0, 1, \ldots, n.$$

Let C_i denote the event $\{\text{chose coin } i\}$ and D_k denote the event that we get k heads from the n tosses. Then $\mathbb{P}C_i = \theta_i$ and

$$\mathbb{P}(D_k \mid C_i) = \binom{n}{k} p_i^k (1 - p_i)^{n-k} \qquad \text{for } k = 0, 1, \ldots, n.$$

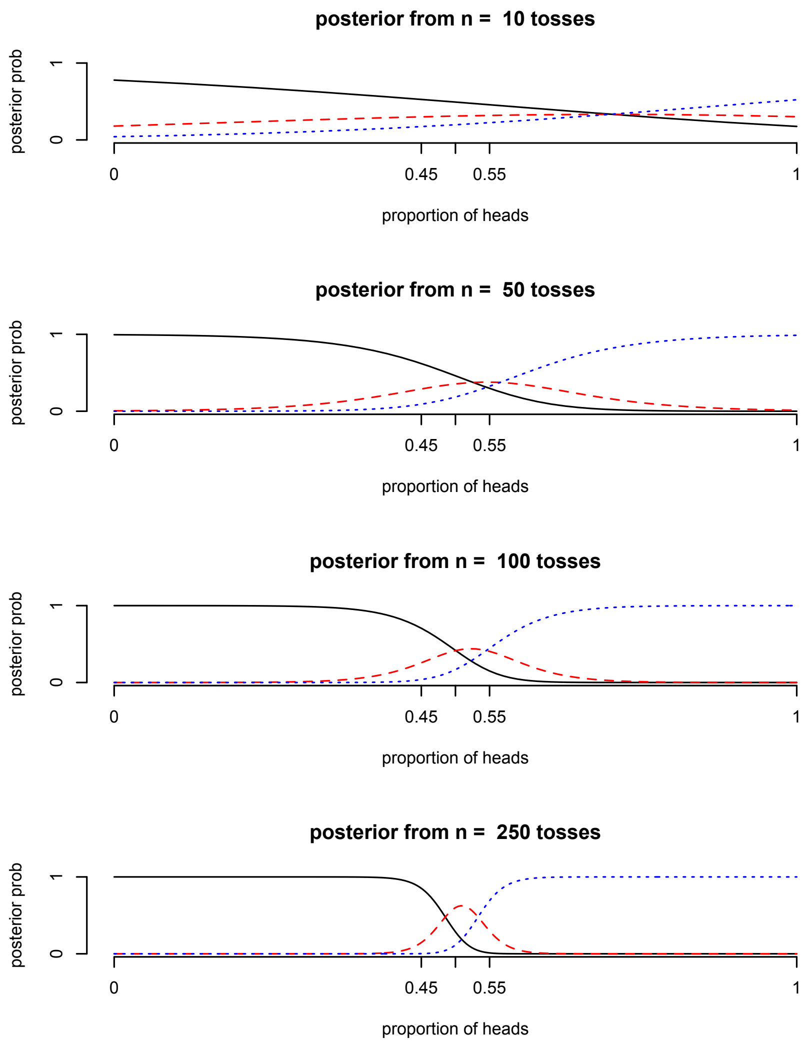
Condition.

$$\begin{aligned} \mathbb{P}(C_i \mid D_k) &= \frac{\mathbb{P}(C_i D_k)}{\mathbb{P}D_k} \\ &= \frac{\mathbb{P}(D_k \mid C_i)\mathbb{P}(C_i)}{\sum_{j=1}^3 \mathbb{P}(D_k \mid C_j)\mathbb{P}(C_j)} \\ &= \frac{p_i^k (1 - p_i)^{n-k} \theta_i}{\sum_{j=1}^3 p_j^k (1 - p_j)^{n-k} \theta_j} \end{aligned}$$

Notice that the $\binom{n}{k}$ factors have cancelled. In fact, we would get the same posterior probabilities if we conditioned on any particular pattern of k heads and $n - k$ tails.

The R-script Bayes.R defines functions to plot the posterior probabilities as a function of k/n, for various choices of the p_i's and the θ_i's and n. The $\mathbb{P}(C_1 \mid D_k)$ are in solid black, the $\mathbb{P}(C_2 \mid D_k)$ are in dashed red, and the $\mathbb{P}(C_3 \mid D_k)$ are in dotted blue. For the pictures I chose $p_1 = 0.45$, $p_2 = 0.5$ and $p_3 = 0.55$ with prior probabilities $\theta_1 = 0.5$, $\theta_2 = 0.3$, and $\theta_3 = 0.2$. The pictures were produced by running:

```
draw.posterior(
        p = c(0.45,0.5,0.55),
        tosses=c(10,50,100,250),
        prior = c(0.5,0.3,0.2)
        )
```



When n gets large, the posterior probability $\mathbb{P}(C_i \mid D_k)$ gets closer to 1 for values of k/n close to p_i, a comforting fact. □