Bernoulli and binomial distributions

David J. H. Shih

Intended learning outcomes

- Recognize and apply Bernoulli and binomial distributions in statistical models
- Derive the binomial and negative binomial distributions

Bernoulli distribution

Definition

A random variable X with domain $\mathcal{X} = \{0, 1\}$ has a Bernoulli distribution if

$$P\left(X = 1 \mid \theta\right) \;=\; \theta$$
$$P\left(X = 0 \mid \theta\right) \;=\; 1 - \theta$$

where $\theta \in [0, 1]$ is the probability of success.

A more compact representation is

$$P\left(X = x \mid \theta\right) \;=\; \theta^{x} \left(1 - \theta\right)^{1-x}, \quad x \in \{0, 1\}$$

We can also write

$$p\left(X = x \mid \theta\right) = \text{Bernoulli}(x \mid \theta)$$
$$X \sim \text{Bernoulli}(\theta)$$

Bernoulli trials

Suppose we have N iid trials with success probability θ, resulting in random variables $X_1, \ldots, X_N$ such that

$$X_i \sim \text{Bernoulli}(\theta).$$

This sequence of iid trials are called *Bernoulli* trials.

Define $\boldsymbol{X} = [X_1, \ldots, X_N]$.

$$\begin{aligned} P_{\boldsymbol{X}}(\boldsymbol{x}) &= \prod_{i=1}^{N} P_{X_i}(x_i) \quad \text{(independent)} \\ &= \prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{\sum_i x_i}(1-\theta)^{1-\sum_i x_i} \end{aligned}$$

Define $Y = \sum_{i=1}^{N} X_i$. Then, for $y \in \{0, 1, \ldots, N\}$,

$$\begin{aligned} P_Y(y) = P\left(\left\{\boldsymbol{x} : \sum_i x_i = y\right\}\right) &= \binom{N}{y} P_{\boldsymbol{X}}(\boldsymbol{x}) \\ = \binom{N}{y} \theta^y (1-\theta)^{N-y}. \end{aligned}$$

Recall from previously,

y	$E = \{\boldsymbol{x} \in \{0,1\}^3 : \sum_i x_i = y\}$
0	$\{000\}$
1	$\{100, 010, 001\}$
2	$\{110, 101, 011\}$
3	$\{111\}$

We have thus derived the **binomial distribution**.

Binomial distribution

Definition

A random variable X has a binomial distribution if

$$P\left(X = x \mid N, \theta\right) \;=\; \binom{N}{x}\theta^{x}\left(1-\theta\right)^{N-x}, \quad x \in \{0, 1, \ldots, N\}$$

where $\theta \in [0, 1]$ is the probability of success.

Binomial theorem

Theorem

For any real numbers x and y, and integer $N \geq 0$,

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

Proof

Base case $N = 0$.
LHS: $(x+y)^0 = 1$.
RHS: $\binom{0}{0} x^0 y^{0-0} = 1$.

Next, we will assume that the equation holds for $N-1$, and prove that it also holds for N (a.k.a. inductive step).

Assume $(x+y)^{N-1} = \sum_{i=0}^{N-1} \binom{N-1}{i} x^i y^{N-1-i}$. Then,

$$\begin{aligned}
(x+y)^N &= (x+y)(x+y)^{N-1} \\
&= (x+y) \sum_{i=0}^{N-1} \binom{N-1}{i} x^i y^{N-1-i} \\
&= \left(\sum_{i=0}^{N-1} \binom{N-1}{i} x^{i+1} y^{N-1-i} \right) + \left(\sum_{i=0}^{N-1} \binom{N-1}{i} x^i y^{N-i} \right)
\end{aligned}$$

In the first sum, re-index with $j = i + 1$. In the second sum, substitute $j = i$. This gives

$$(x+y)^N = \left(\sum_{j=1}^{N} \binom{N-1}{j-1} x^j y^{N-j} \right) + \left(\sum_{j=0}^{N-1} \binom{N-1}{j} x^j y^{N-j} \right)$$

Next, separate out the last term from the first sum, and the first term from the second sum.

$$(x+y)^N = \left(\sum_{j=1}^{N-1} \binom{N-1}{j-1} x^j y^{N-j}\right) + \left(\sum_{j=1}^{N-1} \binom{N-1}{j} x^j y^{N-j}\right)$$
$$+ \binom{N-1}{N-1} x^N y^{N-N} + \binom{N}{0} x^0 y^N$$

By the binomial coefficient identity,

$$
\begin{aligned}
(x+y)^N &= \left(\sum_{j=1}^{N-1} \binom{N}{j} x^j y^{N-j}\right) + \binom{N}{N} x^N y^{N-N} + \binom{N}{0} x^0 y^N \\
&= \sum_{j=0}^{N} \binom{N}{j} x^j y^{N-j}
\end{aligned}
$$

Therefore, by induction, for $N = 0, 1, ...,$

$$
(x+y)^N = \sum_{j=0}^{N} \binom{N}{j} x^j y^{N-j}. \quad \blacksquare
$$

Proof: Binomial(N, θ) is a probability distribution

For $x \in \{0, 1, \dots, N\}$,

$$\text{Binomial}(x \mid N, \theta) = \binom{N}{x} \theta^x (1-\theta)^{N-x}.$$

Non-negativity

$\binom{N}{x} \geq 0$ and $\theta \geq 0$. $1 - \theta \geq 0$ since $\theta \leq 1$.
Therefore, Binomial$(x \mid N, \theta) \geq 0 \quad \forall x$.

Unit measure

For $x = \theta$ and $y = 1 - \theta$, by the binomial theorem,

$$\sum_{x=0}^{N} \binom{N}{x} \theta^x (1-\theta)^{N-x} = (\theta + (1-\theta))^N = 1.$$

Negative binomial distribution

The binomial distribution arises from a sequence of Bernoulli trials where the number of trials N is fixed.

A different distribution can also arise from a sequence of Bernoulli trials if we pre-specify and fix the number of successes R that we wish to observe before we stop.

A random variable X has a negative binomial distribution if

$$P\left(X = x \mid R, \theta\right) \; = \; \binom{x-1}{R-1} \theta^{R} (1-\theta)^{x-R}, \quad x \in \{R, R+1, ...\},$$

where X represents the total number of trials required to achieve a fixed number of successes, $R > 0$.

Illustration

Derivation of the negative binomial distribution

We have X iid trials of success probability θ that contains R successes. X is a random integer variable, whereas $R > 0$ is a fixed integer, and $X \leq R$.

Denote the outcome of each trial by Z_i, so $R = \sum_{i=1}^{X} Z_i$.

Let $\boldsymbol{Z} = [Z_1, Z_2, \dots, Z_X]$ represent the sequence of outcomes. Similarly as before,

$$P_{\boldsymbol{Z}}(\boldsymbol{z}) = \prod_{i=1}^{X} \theta^{z_i}(1-\theta)^{1-z_i} = \theta^{\sum_i z_i}(1-\theta)^{X-\sum_i z_i} = \theta^R(1-\theta)^{X-R}$$

Denote the number of times that each event $\{X = x\}$ occurs as $b(x, R)$, we have

$$P_X(x) = P\left(\{\boldsymbol{z} : \sum_i^x z_i = R\}\right) = b(x, R)\, P_{\boldsymbol{Z}}\left(\boldsymbol{z}\right).$$

For example, if $R = 2$,

x	$E = \{\boldsymbol{z} \in \{0,1\}^x : \sum_i z_i = R\}$
2	$\{11\}$
3	$\{011, 101\}$
4	$\{0011, 0101, 1001\}$
5	$\{00011, 00101, 01001, 10001\}$
...	...

Now, we need to find $b(x, R)$. We know that the xth trial must be a success because we stop when we reach R successes. At trial $x-1$, we have $R-1$ successes, and there are $\binom{x-1}{R-1}$ combinations. Therefore, $b(x, R) = \binom{x-1}{R-1}$, and

$$P_X(x \mid R, \theta) = \binom{x-1}{R-1}\theta^R(1-\theta)^{x-R}. \quad \blacksquare$$

Geometric distribution

A geometric distribution special case of the negative binomial distribution where $R = 1$.

A random variable X has a geometric distribution if

$$P\left(X = x \mid \theta\right) \; = \; \theta(1-\theta)^{x-1}, \quad x \in \{1, 2, ...\},$$

where θ is the probability of success, and X represents the total number trials required for one success.

Summary

"A journey of a thousand miles begins with a single step."
- Lao Tzu

Casella & Berger 2002, sections 3.2, pages 86-91, 95-98.

Intended learning outcomes

- Recognize and apply Bernoulli and binomial distributions in statistical models
- Derive the binomial and negative binomial distributions