The Binomial Distribution

Binomial Experiment

An experiment with these characteristics:

1. For some predetermined number n, a sequence of n smaller experiments called *trials*;
2. Each trial has two outcomes which we call *success* (S) and *failure* (F);
3. Trials are independent;
4. The probability of success is the same in each trial; we denote it $p = P(S)$.

Examples of Binomial Experiments

- Toss a coin or a thumbtack a fixed number of times; S could be "heads" or "lands point up".
- Sample 40 people from a large population; S could be "born in January".

Sampling Without Replacement

A bowl contains 5 red M&Ms and 5 green M&Ms. Choose one at random: $P(\text{red}) = 1/2$.
Without replacing it, choose another:

$$P(\text{red}) = \begin{cases} \frac{4}{9} & \text{if first was red} \\ \frac{5}{9} & \text{if first was green.} \end{cases}$$

Not independent, so *not* a binomial experiment.

Binomial Random Variable

X = number of successes in a binomial experiment

Binomial Distribution

Each outcome of a binomial experiment can be written as a string of n letters, each S or F. The sample space $\mathcal{S}$ is the set of all such strings.

The event $X = x$ is the subset of strings with exactly x Ss, and therefore $(n - x)$ Fs.

Each such string has probability $p^x(1-p)^{n-x}$, and there are $\binom{n}{x}$ of them, so $P(X = x) = \binom{n}{x}p^x(1-p)^{n-x}$.

Shorthand: if X is a binomial random variable with n trials and success probability p, we write $X \sim B(n, p)$.

The range of X is $\{0, 1, \ldots, n\}$. We have shown that the pmf of X, denoted $b(x; n, p)$, is

$$b(x; n, p) = P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, 1, \ldots, n.$$

In R

```
n <- 10
p <- 0.25
probs <-  dbinom(0:n, n, p)
plot(0:n, probs, type = "h")
points(0:n, probs)
```

Expected Value and Variance of X

If $X \sim \text{Bin}(n, p)$ and $q = 1 - p$, then

$$E(X) = np,$$
$$V(X) = np(1-p) = npq,$$

and

$$\sigma_X = \sqrt{npq}.$$

Note that if $n = 1$, X is a Bernoulli random variable; in that case, we showed earlier that $E(X) = p$ and $V(X) = pq$.

Hypergeometric and Negative Binomial Distributions

Hypergeometric Distribution

As noted earlier, sampling *without replacement* from a finite population is *not* a binomial experiment.

A random sample of size n is drawn without replacement from a finite population of N items, of which M are labeled “Success” and the remaining $N - M$ are labeled “Failure”.

Then X, the number of successes in the sample, has the *hypergeometric distribution* with pmf

$$h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

for $\max(0, n - N + M) \leq x \leq \min(n, M)$.

Binomial Approximation

If N is much larger than n, sampling with and without replacement are roughly the same, and the hypergeometric distribution is then close to the binomial distribution:

$$h(x; n, M, N) \approx b\left(x; n, \frac{M}{N}\right)$$

Negative Binomial Distribution

As in a binomial experiment, we carry out a sequence of independent trials, but instead of carrying out a fixed number of *trials*, we wait until we see a fixed number r of *successes*.

Then X, the number of failures that occurred before stopping, has the *negative binomial* distribution with pmf

$$b_{\text{neg}}(x; r, p) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x, x = 0, 1, 2, \ldots$$

Generalized Negative Binomial Distribution

$b_{\text{neg}}(x; r, p)$ can also be defined for non-integer r.

The Poisson Distribution

The binomial distribution, and its special case the Bernoulli distribution, are two of the most important discrete distributions.

The *Poisson distribution* is arguably as important.

Poisson distribution

The random variable X has the Poisson distribution with parameter $\mu > 0$ if its pmf is

$$p(x; \mu) = e^{-\mu} \frac{\mu^x}{x!}, x = 0, 1, 2, \ldots$$

An Approximation

The binomial distribution $B(n, p)$, if n is large and p is small, is approximately the same as the Poisson distribution with parameter $\mu = np$.

To be precise, if $n \to \infty$ and $p \to 0$ in such a way that $np \to \mu > 0$, then

$$b(x; n, p) \to p(x; \mu).$$

Devore's Rule of Thumb

The approximation can safely be used when $n > 50$ and $np < 5$.

What does this mean?

$$\max_{0 \leq x \leq 50} |b(x; 50, 0.1) - p(x, 5)| = .0095 < .01$$

Perhaps that the difference is always less than .01? This table shows the smallest n that achieves this for various values of $\mu = np$.

$\mu = np$	1	2	3	4	5	6	7	8	9	10
Smallest n	19	29	36	42	48	53	58	62	66	70

But for large μ, *all* the individual probabilities are small.

Perhaps we should look at the cdfs:

$$\max_{0 \le x \le 50} |F_b(x; 50, 0.1) - F_p(x, 5)| = .0147 < .015$$

Perhaps the rule of thumb is that the difference in the cdfs is always less than .015? This table shows the smallest *n* that achieves this for various values of $\mu = np$.

$\mu = np$	1	3	5	7	10	15	20	25	30	40
Smallest n	13	31	50	67	95	140	183	228	271	360

Mean and Variance

The Poisson distribution has the interesting property that both its mean and variance are equal to the parameter μ:
If X has the Poisson distribution with parameter μ, then

$$E(X) = V(X) = \mu.$$

For the binomial distribution,

$$V(X) < E(X),$$

and for the negative binomial distribution,

$$V(X) > E(X).$$

Poisson Process

The Poisson distribution is associated with the *Poisson Process*, which is a model for a sequence of events that occur "at random", without memory.

If the events occur in a Poisson Process with *rate* α (events per unit time), then the number of events in any time interval of length t has the Poisson distribution with parameter αt.

The numbers of events in disjoint time intervals are statistically independent.

Bus Arrivals

Suppose you arrive at a bus stop, waiting for a bus that runs every 20 minutes.

If you know nothing about the schedule, the chance that a bus arrives in the next minute is 1 minute/20 $\text{minute}^{-1} = .05$.

Suppose you are told that the last bus arrived 19 minutes ago.

- If the buses run "on time", the chance that a bus arrives in the next minute is now 1.
- But if the schedule slips, it's less than certain.
- If the buses are delayed badly, the information about the last bus is irrelevant, and the buses arrive in a Poisson Process with rate $1/20 = 0.05$ buses per minute.

Radioactive Decay

A Geiger counter detects the emission of nuclear radiation, for instance from a sample of radioactive material.

Emissions are generally assumed to follow a Poisson Process.

Probability Generating Function

The probability generating function (pgf) is a useful tool for studying discrete probability distributions.

If X is a discrete random variable taking non-negative integer values, its pgf is

$$G(z) = E\left(z^X\right) = \sum_{x=0}^{\infty} p(x) z^x.$$

Properties of the pgf

Set $z = 1$:

$$G(1) = E\left(1^X\right) = \sum_{x=0}^{\infty} p(x) = 1.$$

Differentiate and set $z = 1$:

$$G'(z) = \sum_{x=0}^{\infty} p(x) x z^{x-1},$$

so

$$G'(1) = \sum_{x=0}^{\infty} p(x) x = \mu_X.$$

Binomial Distribution

Suppose that $X \sim B(n, p)$; that is, X is the number of successes in n independent trials, with probability p of success in each trial.

For $i = 1, 2, \ldots, n$, write

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{if it is a failure.} \end{cases}$$

Then

$$X = \sum_{i=1}^{n} X_i.$$

So

$$G(z) = E\left(z^X\right) = E\left(z^{\sum X_i}\right) = E\left(\prod_{i=1}^{n} z^{X_i}\right)$$

Because the trials are independent,

$$E\left(\prod_{i=1}^{n} z^{X_i}\right) = \prod_{i=1}^{n} E\left(z^{X_i}\right).$$

Because each X_i is Bernoulli,

$$E\left(z^{X_i}\right) = z^0(1-p) + z^1 p = 1 - p + pz = q + pz,$$

and so

$$G(z) = (q + pz)^n.$$

By the binomial theorem,

$$(q + pz)^n = \sum_{x=0}^{n} \binom{n}{x} (pz)^x q^{n-x},$$

so

$$b(x; n, p) = \text{coefficient of } z^x = \binom{n}{x} p^x q^{n-x}, 0 \leq x \leq n.$$

Poisson Approximation

Suppose that n is large, and $p = \mu/n$ for some $\mu > 0$. Then

$$\begin{aligned} G(z) &= \left(1 - \frac{\mu}{n} + \frac{\mu}{n} z\right)^n \\ &= \left(1 + \frac{\mu(z-1)}{n}\right)^n \\ &\approx e^{\mu(z-1)}. \end{aligned}$$

Now

$$e^{\mu(z-1)} = e^{-\mu} e^{\mu z} = e^{-\mu} \sum_{x=0}^{\infty} \frac{(\mu z)^x}{x!}$$

so

$$b(x; n, p) \approx e^{-\mu} \frac{\mu^x}{x!},$$

the pmf of the Poisson distribution.