STAT 315: Binomial Random Variables

Luc Rey-Bellet

University of Massachusetts Amherst

luc@math.umass.edu

February 27, 2025

Binomial trials

Binomial experiments

- The experiment consists of n identical trials.
- Each trial results in one of two outcomes: success S or failure F.
- Each trial has probability of success p and probability of failure $q = (1 - p)$.
- The trials are independent.
- The binomial random variable Y is

$$Y = \text{ number of successes observed during the } n \text{ trials}$$

The random variable Y has two parameters
$n =$ numbers of trials
$p =$ probability of success
Notation: We write $Y \sim B_{n,p}$

Math reminder: The binomial theorem

Theorem

The binomial theorem

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

Proof:

$$(x+y)^n = \underbrace{(x+y)(x+y)\cdots(x+y)}_{n \text{ times}}$$

Expanding the product gives a sum of 2^n terms. Each term has the form $x^k y^{n-k}$ with $k = 0, 1, \cdots, n$.

The number of terms of the form $x^k y^{n-k}$ is equal to $\binom{n}{k}$ since it is the number of ways to choose the k locations for the x.

Adding all the terms up give the theorem.

The PDF of a binomial random variable

Probability distribution function

- Y takes values $0, 1, 2, \cdots, n$ (the number of successes).
- The pdf is

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

For example with $n = 5$ a sequence of 2 failures and three successes e.g. $FSSFS$ has probability $p^3(1-p)^2$ and there are $\binom{5}{3}$ such sequences (=number of ways to choose which ones of the five trials is a success).

Note that by the binomial theorem

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

and so $p(k)$ is a pdf!

Shape of the binomial distribution

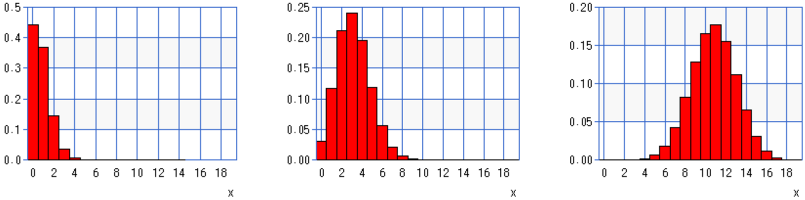

Figure: The binomial pdf for $B_{20,.04}$, $B_{20,.16}$, and $B_{20,.53}$ from left to right.

$$\frac{p(k)}{p(k-1)} = \frac{\frac{n!}{(k)!(n-k)!} p^k (1-p)^{n-k}}{\frac{n!}{(k-1)!(n-k+1)!} p^{k-1}(1-p)^{n-k+1}} = \frac{n-k+1}{k} \frac{p}{1-p}$$

and thus $\dfrac{p(k)}{p(k-1)} \geq 1 \text{ iff } (n-k+1)p > k(1-p) \text{ iff } (n+1)p \geq k$

The maximum is around $k_{max} \approx (n+1)p$.

Example

- A multiple-choice examination has 10 questions, each with four possible answers, only one of which is correct. Having not studied a student answers each of the questions with an independent random guess.
  - What is the probability he get 2 or 3 questions correct?
  - What is the probability the students gets a passing score (at least 6 section correct)?

Chuck-a-luck (USA) or Bâu cua tôm cá (gourd-crab-shrimp-fish, Vietnam) or Hoo Hey How (Fish-Prawn-Crab, China)




The game is played with 3 six-faced dice a board with the pictures (or numbers) on the six faces.

If you bet \$1 on say "shrimp", you win \$k if you land k "shrimps", $k = 1, 2, 3$ and lose your \$1 otherwise.

Mean and Variance of the binomial RV

Mean and Variance of a binomial RV $Y \sim B_{n,p}$

If n is the number of trials and p the probability of success

$$E[Y] = \mu = np$$
$$V[Y] = \sigma^2 = np(1-p)$$

Note that $E[Y]$ is order n and so maybe one would guess that $V[Y] = E[Y^2] - E[Y]^2$ should be of order n^2. But this is not so (some cancellation occurs..)

Note the following identity

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

Pick a team of k then a captain versus Pick a captain and then the rest of the team.

$$\begin{aligned} E[Y] &= \sum_{k=0}^{n} k\binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} n\binom{n-1}{k-1} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1}(1-p)^{n-k} = np \underbrace{\sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{n-1-j}}_{=1} \\ &= np \end{aligned}$$

Similarly

$$k(k-1)\binom{n}{k} = \frac{n!}{(k-2)!(n-k)!} = n(n-1)\binom{n-2}{k-2}$$

and

$$\begin{aligned} E[Y(Y-1)] &= \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k (1-p)^{n-k} \\ &= n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2}(1-p)^{n-k} \\ &= n(n-1)p^2 \underbrace{\sum_{j=0}^{n-2} \binom{n-2}{j} p^j (1-p)^{n-2-j}}_{=1} = n(n-1)p^2 \end{aligned}$$

So

$$\begin{aligned} V(Y) &= E[Y^2] - E[Y]^2 = E[Y(Y-1)] + E[Y] - E[Y]^2 \\ &= n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p) \end{aligned}$$

Application: Pool testing

For example suppose we have 50 samples to test for COVID 19 and tests are expensive.

- Option 1: Test everyone (50 tests).
- Option 2:
  1. Split 50 samples into 10 pools of 5.
  2. Test 10 pools (=10 tests).
  3. If a pool is positive then retest everyone in the pool separately to identify the sick ones

  For example if 4 pools out 10 test positive did $10 + 20 < 50$ tests.

How should we pick the pool size?

How does it depend on the rate of infection?

Given

- $N =$ size of the population.
- $k=$ size of the pool.
- $p =$ probability to be infected.

Let us compute the expected number of test.
In the first round we perform $\frac{N}{k}$ tests, one for each pool of size k.
In the second round we have (assuming the tests are perfect)

$$P(\text{one pool test negative}) = P(\text{everyone is negative}) = (1-p)^k$$

so

$$P(\text{one pool tests positive}) = 1-(1-p)^k$$

So the number of positive pools is a binomial random variable with number of trials $n = \frac{N}{k}$ and probability of "success" equal to $1-(1-p)^k$. So

$$\text{Expected number of pools needed retesting} = k \times \frac{N}{k}\left(1-(1-p)^k\right)$$

$$\text{Expected number of tests} = \frac{N}{k} + N\left(1 - (1-p)^k\right) = N\left(\frac{1}{k} + (1-p)^k\right)$$

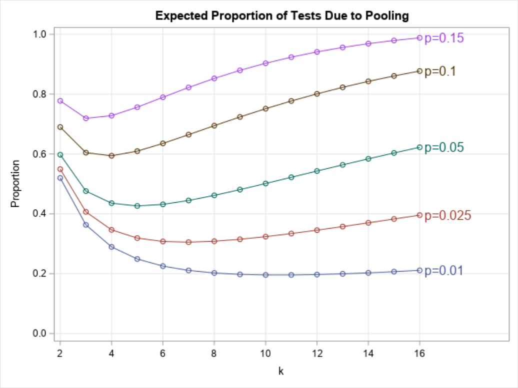

Figure: plot of $1/k + (1-p)^k$

Find the minimum of $1/k + (1-p)^k$.

For example for $p = 0.1$ the minimum is $k = 4$ so one should make group of 4 to be tested together and one saves 40% on test costs.