

Discrete Distributions: A Round-up

Bernoulli trials

Binomial $X \sim \mathcal{B}(n, p)$: number of successes in n trials

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

Geometric $X \sim G(p)$: number of trials TO GET the first success

$$P(X = k) = (1 - p)^{k-1} p, \quad k \geq 1.$$

Negative Binomial $X \sim \mathcal{NB}(p, m)$: number of trials TO GET m successes; $X = m, m + 1, \dots$

$$P(X = k) = \binom{k-1}{m-1} p^{m-1} (1-p)^{(k-1)-(m-1)} p = \binom{k-1}{m-1} p^m (1-p)^{(k-m)}$$

($m - 1$ places to place the successes in the first $k - 1$ trials; k -th trial has to be a success)

Runs: in 64 tosses of a fair coin, 6 consecutive heads or 6 consecutive tails are likely.



Jacob (Jacques) Bernoulli (1654 – 1705): Swiss.

Some contributions:

- First law of large numbers: $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = p$, if X_k are

independent $\mathcal{B}(1, p)$;

- Bernoulli differential equation $y' = p(x)y + q(x)y^n$;
- polar coordinates;
- Brother of Johann Bernoulli and uncle of Daniel Bernoulli; his own two children were not mathematicians.

Poisson Distribution

$$X \sim \mathcal{P}(\lambda) : P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots; E(X) = \text{Var}(X) = \lambda.$$

Poisson approximation of binomial distribution:

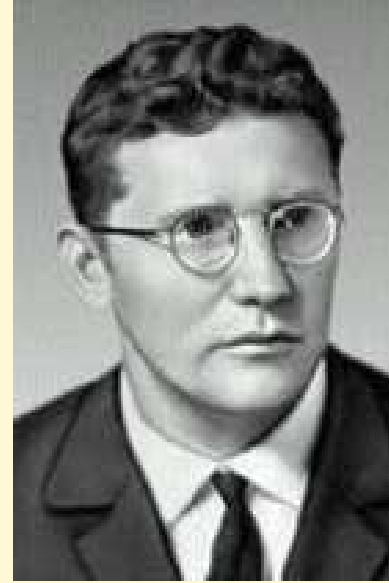
$$\lim_{n \rightarrow \infty, p \rightarrow 0, np \rightarrow \lambda} \binom{n}{k} p^k (1-p)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Quality of Approximation:

$$\sum_{k=0}^{\infty} \left| \binom{n}{k} p^k (1-p)^{n-k} - \frac{\lambda^k}{k!} e^{-\lambda} \right| \leq \frac{2\lambda}{n} \min(2, \lambda), \quad \lambda = np,$$

a theorem of Prokhorov (for ideas, see Chapter 8 of the book)

Poisson and Prokhorov



Siméon Denis Poisson (1781–1840): French (distribution: 1837).
Yuri Vasilevich Prokhorov (1929–2013): Russian.

Hyper-Geometric Distribution: Definition

The main thing to remember: 23 balls; 11 red, 12 black; take 9;

$$P(5 \text{ red}) = \frac{\binom{11}{5} \binom{12}{4}}{\binom{23}{9}}$$

General description: $X \sim \mathcal{H}(N, m; n)$

N objects of two types; m of type I; n is sample size; k is the number of Type I in the sample.

$$P(X = k) = \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \dots, \min(m, n).$$

Key words: without replacement.

From Hyper-Geometric to Binomial

Theorem

$$\lim_{N \rightarrow \infty, m \rightarrow \infty, m/N \rightarrow p} \frac{\binom{m}{k} \binom{N-m}{n-k}}{\binom{N}{n}} = \binom{n}{k} p^k (1-p)^{n-k}$$

Then can follow up with a Poisson approximation.

Example: a raffle; n^2 tickets, n winning. You buy $2n$ tickets.

$$P(3 \text{ win}) = P(\mathcal{H}(n^2, n; 2n) = 3) \approx P(\mathcal{B}(2n, n^{-1}) = 3) \approx P(\mathcal{P}(2) = 3).$$

does not depend on n (exact in the limit $n \rightarrow \infty$).

Bottom line: $\mathcal{H}(N, m; n) \approx \mathcal{B}(n, m/N) \approx \mathcal{P}(nm/N)$

zeta distribution:

$$P(X = k) = \frac{1}{\zeta(\alpha + 1)k^{1+\alpha}}, \quad \alpha > 0, \quad k = 1, 2, \dots$$

Choosing a suitable α can fit many models...

Riemann zeta function: $\zeta(t) = \sum_{k=1}^{\infty} \frac{1}{k^t}, \quad t > 1.$

Zipf distribution: $P(X = k) = C(N)k^{-1}, \quad k = 1, \dots, N.$

Frequency of N ranked objects.

Erlang's distribution:

$$P(X = k) = C(N) \frac{\lambda^k}{k!}, \quad k = 0, \dots, N.$$

(truncated Poisson distribution)

Who are those people?



Library of Cong



Georg Friedrich Bernhard Riemann (1826–1866): German (1859)

George Kingsley Zipf (1902–1950): American (1932)

Agner Krarup Erlang (1878–1929): Danish (1906)