Lecture Summary

5.2 The Bernoulli and Binomial Distributions

- SKIP: [5.3] The Hypergeometric Distributions

5.4 The Poisson Distributions

Bernoulli distributions

Definition (Bernoulli distribution)

A r.v. X has the Bernoulli distribution with parameter p if $P(X = 1) = p$ and $P(X = 0) = 1 - p$. The probability function (pf) of X is

$$f(x|p) = \begin{cases} p^x(1-p)^{1-x} & x = 0, 1 \\ 0 & otherwise \end{cases}$$

- An experiment with two outcomes: "success", "failure", X = number of successes.
- Parameter space: $p \in [0, 1]$.
- $E(X) = p, Var(X) = p(1-p)$.
- MGF: $\psi(t) = E(e^{tX}) = pe^t + (1-p)$.
- CDF:

$$F(x) = \begin{cases} 0, & x < 0 \\ 1-p, & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Binomial distributions

A r.v. X has the Binomial distribution with parameters n and p a if the probability function (pf) of X is

$$f(x|p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \dots, n \\ 0 & \textit{otherwise} \end{cases}$$

- n repetitions of an experiment with two outcomes: "success", "failure", X = number of successes.
- Parameter space: n positive integer, $p \in [0, 1]$.
- $E(X) = np$, $Var(X) = np(1-p)$.
- CDF:

$$F(x) = \begin{cases} 0, & x < 0 \\ 1-p, & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Poisson Distributions

Definition (Poisson Distributions)

Let $\lambda > 0$. A random variable X follows the *Poisson distribution* with mean λ if the p.m.f. of X is as follows:

$$f(x|\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & x = 0, 1, 2, \ldots \\ 0 & \textit{otherwise}. \end{cases}$$

- Parameter space: $\lambda \in [0, \infty)$.
- $E(X) = \lambda$, $Var(X) = \lambda$

Poisson Distributions

The Poisson distribution is useful for modeling uncertainty in counts / arrivals.

Examples:

- How many calls arrive at a switch board in one hour?
- How many busses pass while you wait at the bus stop for 10 min?
- How many customers will enter a store in 15 minutes?

Properties of the Poisson

Theorem (Sum of Poissons is a Poisson.)

Let $X_1, \ldots X_k$ are independent and if X_i has the Poisson distribution with mean $\lambda_i (i = 1, \ldots, k)$, then the sum $X_1 + \cdots + X_k$ has the Poisson distribution with mean $\lambda_1 + \cdots + \lambda_k$.

Theorem (Approximation to the Binomial)

For each integer n and each $0 < p < 1$, let $f(x|n, p)$ denote the pf of the Binomial distribution with parameters n and p, and let $f(x|\lambda)$ denote the pf of the Poisson distribution with mean λ. Let $\{p_n\}_1^\infty$ be a sequence of numbers between 0 and 1 such that $lim_{n \to \infty} = \lambda$. Then

$$lim_{n \to \infty} f_{X_n}(x|n, p_n) = f(x|\lambda)$$

When the value of n is large, and the value of p is very small, the Poisson with mean np is a good approximation for the Binomial with parameters n and p.

Question

- We need to test 1000 people for a rare disease (affects two in 1000 people).
- Each test requires a small amount of blood, and it is guaranteed to detect the disease if it is anywhere in the blood.
- Strategy 1: Test 1000 people
- Strategy 2: Split in 10 groups of 100, combine their bloods and test them. (10 tests). If any of them tests positive, test all people in that group.
- What is the expected number of tests for Strategy 2?

Question

Hints

- Let Z_i be the number of people in group i that have the disease, $i = 1, \ldots, 10$.
- What is the distribution of each Z_i?
- Let $Y_i = 0$ if group i tests positive, $Y_i = 1$ otherwise.
- What is the distribution of Y_i?
- Let $Y = Y_1 + \cdots + Y_{10}$ be the number of groups where every individual has to be tested.
- What is the distribution of Y?
- How many tests are we going to need as a function of the value of Y?
- What is the expected value of the number of tests we will need?

Question

Hints (1)

- Let Z_i be the number of people in group i that have the disease, $i = 1, \dots, 10$.
- What is the distribution of each Z_i?
  - $Z_i \sim Binomial(100, 0.002)$
- Let $Y_i = 1$ if group i tests positive, $Y_i = 0$ otherwise.
- What is the distribution of Y_i?
  - Y_i follows a Bernoulli distribution (is either 0 or 1). $P(Y_i) = 1 = P(Z_i > 0) = 1 - P(Z_i = 0) = 1 - (1 - 0.002)^{100} = 0.181$

Question

Hints

- Let $Y = Y_1 + \cdots + Y_{10}$ be the number of groups where every individual has to be tested.
- What is the distribution of Y?
  - $Y \sim Binomial(10, 0.181)$
- How many tests are we going to need as a function of the value of Y?
  - $X = 10 + 100Y$
- What is the expected value of the number of tests we will need?
  - $E(X) = E(10 + 100Y) = 10 + 100E(Y) = 10 + 100 * 10 * 0.181 = 191$