STA2001 Probability and Statistics (I)

Lecture 7

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

Review

Definition[Mathematical Expectation]

Assume that X is a discrete RV with range $\overline{S}$ and $f(x)$ is its pmf. If $\sum_{x\in\overline{S}} g(x)f(x)$ exists, then it's called the mathematical expectation of $g(X)$ and is denoted by

$$E[g(X)] = \sum_{x\in\overline{S}} g(x)f(x)$$

Property[Mathematical Expectation]

Mathematical expectation is a linear operator, i.e.,

$$E[c_1 g_1(X) + c_2 g_2(X)] = c_1 E[g_1(X)] + c_2 E[g_2(X)]$$

Review

Definition[Special mathematical expectation]

$$E[g(X)] = \sum_{x \in \overline{S}} g(x) f(x)$$

$$g(X) = \begin{cases} X \rightarrow \text{ Mean} \\ (X - EX)^2 \rightarrow \text{ Variance} \\ X^r \rightarrow \text{ Moment} \\ e^{tX}, \text{ for } |t| < h, \rightarrow \text{ Mgf:} M(t) = \begin{cases} M(0) = 1 \\ M'(0) = E[X] \\ M''(0) = E[X^2] \end{cases} \end{cases}$$

2.4 Binomial Distribution

Starting from this section, we will study some typical random phenomena/experiments and corresponding distributions, which are described by RV

1. description of the random phenomena/experiments
2. pmf (probability function), cdf
3. mathematical expectations, e.g., mean, variance, mgf

Bernoulli Experiment

Description: The outcomes can be classified in one of two mutually exclusive and exhaustive ways, say either

success or failure

female or male

life or death

Bernoulli Experiment

Description: The outcomes can be classified in one of two mutually exclusive and exhaustive ways, say either

success or failure

female or male

life or death

Bernoulli Distribution

Let X be a RV associated with a Bernoulli experiment with the probability of success p.

- RV: $X : S \to \overline{S}, S = \{\text{success, failure}\}$. Define

  $X(\text{success}) = 1,\ X(\text{failure}) = 0,\ \overline{S} = \{0, 1\}$

- pmf of $X : f(x) : \overline{S} \to [0, 1]$

$$f(x) = p^x(1-p)^{1-x},\ x \in \overline{S}$$

Then we say X has a Bernoulli distribution with probability of success p.

Bernoulli Distribution

Mathematical expectations:

1. $E[X]$
2. $Var[X]$
3. $M(t) = E[e^{tX}]$

Bernoulli Distribution

Mathematical expectations:

1. $E[X] = \sum_{x \in \overline{S}} xf(x) = 0 \cdot (1-p) + 1 \cdot p = p$
2. $$\begin{aligned} Var[X] = E[(X - EX)^2] &= \sum_{x \in \overline{S}} (x-p)^2 f(x) \\ &= p^2(1-p) + (1-p)^2 p = (1-p)p \end{aligned}$$
3. Mgf: $M(t) = E[e^{tX}] = e^t \cdot p + (1-p), \quad t \in (-\infty, \infty)$

Bernoulli Trials

If a Bernoulli experiment is performed n times

1. independently, i.e., all trials are independent
2. the probability of success, say p, remains the same from trial to trial.

then these n repetitions of the Bernoulli experiment is called n Bernoulli trials.

Example 1

For a lottery, the probability of winning is 0.001. If you buy the lottery for 10 successive days, that corresponds to 10 Bernoulli trials with the probability of success $p = 0.001$.

Random sample of size n from a Bernoulli distribution

In a sequence of n Bernoulli trials, let X_i denote the Bernoulli RV associated with the ith trial.

An observed sequence of n Bernoulli trials will be n-tuple of zeros and ones, which is called a random sample of size n from a Bernoulli distribution.

Example 2, page 74

Instant lottery ticket; 20% are winners. 5 tickets are purchased and $(0, 0, 0, 1, 0)$ is a random sample. Assuming independence between purchasing different tickets, What is probability of this sample?

Example 2, page 74

Recall that if all trials are independent and let A_i be the event associated with the ith trial. Then

$$P(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$$

Therefore, the probability is $0.2(0.8)^4$ according to multiplication principle for independent events.

Binomial Distribution

We are interested in the number of successes in n Bernoulli trials. The order of the occurrences is not relevant.

Let X be the number of successes in n Bernoulli trials with its range $\overline{S} = \{0, 1, 2, \cdots, n\}$. Find the pmf of X.

1. A Bernoulli (success-failure) experiment is performed n times.
2. The n trials are independent $P(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$, where A_i is the event associated with ith trial, multiplication rule for independent events.
3. The probability of success for each trial is p.

Binomial Distribution

4. If $x \in \overline{S}$ successes occur, the number of ways of selecting x successes in n Bernoulli trials is $\binom{n}{x} = \frac{n!}{x!(n-x)!}$. Since Bernoulli trials are independent, the probability of each way is $p^x(1-p)^{n-x}$

$$\Rightarrow f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \cdots, n$$

Binomial Distribution

Definition[Binomial distribution]

A RV X is said to have a binomial distribution, if the range space $\overline{S} = \{0, 1, \cdots, n\}$ and the pmf

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \cdots, n$$

and denoted by $X \sim b(n, p)$, where the constants n, p are parameters of the distribution.

It is called the binomial distribution because of its connection with binomial expansion

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} \text{ with } a = p, \quad b = 1-p$$

Example 2 [revisited]

If X is the number of winning tickets among 5 tickets that are purchased. What is the probability of purchasing 2 winning tickets?

Example 2 [revisited]

If X is the number of winning tickets among 5 tickets that are purchased. What is the probability of purchasing 2 winning tickets?

$$X \sim b(5, 0.2), \quad f(2) = P(X = 2) = \binom{5}{2}(0.2)^2(0.8)^3.$$

Mgf of Binomial Distribution

Let $X \sim b(n,p)$. Then by definition,

$$M(t) = E[e^{tX}] = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$
$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x}$$
$$= [(1-p) + pe^t]^n \quad -\infty < t < \infty$$

From the expansion of

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x} \text{ with } a = pe^t, \quad b = 1-p$$

Mgf of Binomial Distribution

Question

What is the use of mgf?

Mgf of Binomial Distribution

Question

What is the use of mgf?

$M(t) = (pe^t + 1-p)^n, \quad t \in R$

$$M'(t) = n[(1-p) + pe^t]^{n-1}pe^t \Rightarrow M'(0) = E[X] = np$$

$$M''(t) = n(n-1)[(1-p) + pe^t]^{n-2}p^2e^{2t} + n[(1-p) + pe^t]^{n-1}pe^t$$

$$M''(0) = E[X^2] = n(n-1)p^2 + np$$

$$Var[X] = E[X^2] - (E[X])^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p)$$

By the way, when $n = 1$ in $b(n, p)$, the binomial distribution reduces to Bernoulli distribution denoted by $b(1, p)$.

cdf of Binomial Distribution

cdf of Binomial Distribution

$$F(x) = P(X \le x) = \sum_{y \in \{X \le x\}} f(y) = \sum_{y=0}^{[x]} \binom{n}{y} p^y (1-p)^{n-y},$$

where $x \in (-\infty, \infty)$ and $[x]$ is the largest integer $\le x$.

Example 3

A kind of chicken are raised for laying eggs. Let $p = 0.5$ be the probability that the newly hacked chick is a female. Assuming independence, let X be the number of female chicken out of 10 newly hatched chicks selected at random.

$$P(X \leq 5)?$$

$$P(X = 6)?$$

$$P(X \geq 6)?$$

Example 3

Then $X \sim b(10, 0.5)$

$$P(X \le 5) = \sum_{x=0}^{5} \binom{10}{x} 0.5^x 0.5^{\overset{10}{\cancel{5}}-x}$$

$$P(X = 6) = \binom{10}{6} 0.5^6 0.5^4 = P(X \le 6) - P(X \le 5)$$

$$P(X \ge 6) = 1 - P(X \le 5)$$