VE401 SU2022 RC week2

Shuyu Wu

UM-SJTU Joint Institute

wushuyu2002@sjtu.edu.cn

May 11, 2022

Outline

Random Variable

definition: A map from sample space to a subset of real number,

$$X : S \to \mathbb{R}$$

It may look strange, and now let's interpret this definition.

Suppose in one round of Bernoulli trial, there's p probability to get result A and $1 - p$ probability to get result B. We record the result A we get in 10 rounds of experiment.

The sample space here is $S = \{A, B\}^{10}$. The possible result is 0, 1, 2, ..., 10. For each result $m \in S$, we define $X(m)$ as the result A contained in m. Here, X is a map from the set of all possible result of the experiment(i.e., the sample space S), to a subset of $\mathbb{R}$.

From this example, we can see that the value of the random variable is closely related to the result of the random experiment. Before random experiment, the value of the random variable is unknown, but have certain probability for every possible value of it.

Random Variable

The meaning of introducing "Random Variable" into probability theory, is to turn our research target from various kind of sample space to real number. We can then apply **mathematical analysis** method to tackle problems, as what you have learnt in VV186 or VV156. This is the most important difference between high school level and university level probability theory.

Random variable can be classified into 2 general types or 3 kinds:

- discrete random variable: the possible value(range) is at most countable.
- non-discrete random variable
  - continuous random variable
  - neither discrete nor continuous random variable

probability density function(PDF)

For discrete random variable X, it has at most countable possible result. Suppose the set of all possible result is Ω, the probability density function is a map

$$f_X : \Omega \to \mathbb{R}$$

and $f_X(x)$ is $P[X = x]$.

Outline

Bernoulli Distribution

Bernoulli(or 0-1) distribution: If a random variable X only have two possible result: 0 and 1. The probability to get 1 is p. We then say that X follows a Bernoulli distribution(or 0-1 distribution) with parameter p, written as $X \sim Bernoulli(p)$

The PDF for Bernoulli distribution is $P[X = k] = p^k(1 - p)^{1-k}(k = 0, 1)$

This is the simplest distribution, but it's the basis of an important distribution: binomial distribution.

Binomial Distribution

Independent and Identical Distributed(i.i.d): Two or more random variables are called independent and identical distributed if they are independent and follow the same distribution.

The inspiration of binomial distribution is the success number of n i.i.d bernoulli distribution, i.e., $P[x$ success in n trials$]$.

The PDF of binomial distribution with parameters n, p is

$$P[X = x] = \binom{n}{x} p^x (1 - p)^{n-x}$$

We have already proved this in last RC slide.

ex 2.1

1. A new recruit is having his shooting training. The probability to hit the target is 0.1. He shoots 20 bullets, what's the probability that he hit the target at least twice?

2. Many people think they can tell Coca Cola and Pepsi apart while they can't. A kind of test is to prepare 2 cup of Coca Cola and one cup of Pepsi and asked the tester to tell which one is different. Now A and B are asked to take this kind of test. Suppose A actually doesn't have the ability to tell Coca Cola and Pepsi apart while B has 0.6 probability to identify whether a specified cup of cola is Coca or Pepsi.
Now they have 3 rounds of test, and suppose A and B don't know whether the different cola is Coca or Pepsi. What's the probability that the success times of B is higher than A?

ex 2.1 answer

1. Set X to be the number he hits the target. $X \sim B(20, 0.1)$ and $P[X = x] = \binom{20}{x} 0.1^x 0.9^{20-x}$, we want to solve $P[X \geq 2] = 1 - P[X = 0] - P[X = 1] = 0.61$

2. For A, he can only test his luck, so he has $\frac{1}{3}$ probability to succeed in each experiment.

Now let's calculate the probability that B succeeds in a round of test. If B happens to judge the type of 3 cups of cola all correct or all wrong, he can tell which one is correct. If he wrongly judges one cup of Pepsi or two cups of Coca he will think three cups are the same, and he has $\frac{1}{3}$ probability to succeed. The total probability for him to succeed in one test is

$$0.6^3 + 0.4^3 + \frac{0.6^2 * 0.4 + 0.4^2 * 0.6}{3} = 0.36$$

ex 2.1 answer

Now, denote X_1 to be the success number of A, and X_2 to be the success number of B in 3 rounds of test. We have $X_1 \sim B(3, \frac{1}{3})$, $X_2 \sim B(3, 0.36)$. So

$$P[X_1 < X_2] = P[X_1 = 0]P[X_2 \geq 1] + P[X_1 = 1]P[X_2 \geq 2] + P[X_1 = 2]P[X_2 = 3]$$

The answer is $\frac{3416 + 2052 + 162}{15625} = 0.36032$

Cumulative Distribution Function(CDF)

For a random variable X, the **Cumulative Distribution Function** of X is defined as

$$F_X(x) = P[X \leq x]$$

Geometric Distribution

The binomial distribution is the success number in n rounds of i.i.d Bernoulli trial. On the other hand, the **geometric distribution** focuses on the total rounds of i.i.d bernoulli trial to get the first success.

Set X to be the total trials to get the first success, and p is the success rate for each trial. Then $P[X = x] = (1 - p)^{x-1}p$
The CDF of geometric distribution is $F_X(x) = 1 - (1 - p)^{[x]}$

Outline

Expectation

definition: Suppose X is a discrete random variable, and the set of the possible result of X is Ω, the expectation of X is defined as

$$E[X] := \sum_{x \in \Omega} x \cdot f_X(x)$$

If Ω is an infinite subset of $\mathbb{R}$, we require the series above to be **absolutely converge**.

Important properties

- $E[X + Y] = E[X] + E[Y]$
- $E[cX] = cE[X]$
- If $Y = \varphi(X)$, then $E[Y] = \sum_{x \in \Omega} \varphi(x) \cdot f_X(x)$

Variance and Standard Deviation

definition: The Variance of X is $Var[X] := E[(X - E[X])^2]$
The standard deviation of X is $\sigma_X := \sqrt{Var[X]}$

Important properties

- $Var[X] = E[X^2] - E[X]^2$
- $Var[cX] = c^2 Var[X]$
- If X and Y are independent, then $Var[X + Y] = Var[X] + Var[Y]$

The expectation and variance of some distribution

- binomial distribution: Suppose $X \sim B(n, p)$
  - $E[X] = np$
  - $Var[X] = np(1 - p)$
- geometric distribution: Suppose $X \sim Geom(p)$
  - $E[X] = \frac{1}{p}$
  - $Var[x] = \frac{1-p}{p^2}$

Moment

The nth (ordinary) moment of X is defined as $E[X^n]$

The nth central moment of X is defined as $E[(\frac{X-E[X]}{\sigma_X})^n]$

Moment-Generating Function(m.g.f)

For a random variable, the nth derivative of its moment generating function is $E[X^n]$.

$m_X(t) = E[e^{tX}]$

- can be used to calculate expectation and variance
- The m.g.f can uniquely determine PDF.
- The m.g.f of binomial distribution is $(q + pe^t)^n, q = 1 - p$
- The m.g.f of geometric distribution is $\frac{pe^t}{1-qe^t}$

ex 2.2

1. Prove that $Var[X + Y] = Var[X] + Var[Y]$ when X and Y are independent.

2. Prove that $m_{X+Y}(t) = m_X(t)m_Y(t)$ when X and Y are independent.(This is useful to determine the distribution of $X + Y$)

ex 2.2 answer

1.

$$Var[X + Y]$$
$$= E[(X + Y)^2] - E[X + Y]^2$$
$$= E[X^2 + Y^2 + 2XY] - (E[X]^2 + E[Y]^2 + 2E[X]E[Y])$$
$$= Var[X] + Var[Y] + 2(E[XY] - E[X]E[Y])$$
$$E[XY] = \sum_{\Omega_1 * \Omega_2} xyP[X = x, Y = y]$$
$$= \sum_{\Omega_1 * \Omega_2} (xP[X = x])(yP[Y = y])$$
$$= E[X]E[Y]$$

So $Var[X + Y] = Var[X] + Var[Y]$

ex 2.2 answer

2.

$$m_X(t) = E[e^{tx}]$$
$$m_Y(t) = E[e^{ty}]$$
$$m_{X+Y}(t) = E[e^{t(X+Y)}]$$
$$= E[e^{tX}e^{tY}]$$
$$= E[e^{tX}]E[e^{tY}] \text{(X and Y are independent)}$$
$$= m_X(t)m_Y(t)$$

With this formula we can find the distribution of $X + Y$ by multiplying the m.g.f of X and Y and see its result.

Outline

Pascal Distribution

Pascal distribution is the generalization of geometric distribution.
Geometric distribution focuses on the total trials towards 1 success, while pascal distribution focuses on the total trials towards r success.
To get r success, let's calculate the probability that we need x trials. We should have r-1 success in x-1 trials and the last trial is success. So, the probability is

$$\binom{x-1}{r-1} p^{r-1} (1-p)^{x-r} p = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

This is the PDF for the Pascal distribution with parameters p and r.
The m.g.f of Pascal distribution is $\left(\frac{pe^t}{1-qe^t}\right)^r$

Negative Binomial Distribution

Negative binomial distribution is very similar to Pascal distribution, except that it record the faliure number instead of total number. The PDF for negative binomial distribution is $\binom{x+r-1}{r-1} p^r (1-p)^x$

Poisson Distribution

(This example is adapted from 2022 SJTU MCM/ICM training camp)

Suppose your restaurant has λ customers on average everyday. You want to estimate the number of customers for the next day to prepare the food more efficiently.

Suppose that the probability for a customer to come is independent of time(which means that the customer number around 7:00 is roughly the same as the number at 12:00 or 15:00). What's the probability distribution of next day's customers?

Poisson Distribution

Suppose one day is uniformly seperated into n very small time interval. Each time interval is $\frac{1}{n}$ day, which is so small that at most one customer come in one interval. So, the probability that a customer come in an interval is $\frac{\lambda}{n}$. We need to notice that whether a customer come in each interval is independent event.

Dennote X to be the total customers come this day, which follows a binomial distribution $B(n, k)$. That is to say,

$$P[X = x] = \binom{n}{x} p^x (1-p)^{n-x} \quad (p = \frac{\lambda}{n})$$

$$= \binom{n}{x} (\frac{\lambda}{n})^x (\frac{n-\lambda}{n})^{n-x}$$

$$= \frac{n!}{x!(n-x)!} (\frac{\lambda}{n})^x (\frac{n-\lambda}{n})^{n-x}$$

Take limit $n \to \infty$ and we get:

Poisson Distribution

$$P[X = x] = \frac{\lambda^x e^{-\lambda}}{x!}$$

This is the PDF of Poisson distribution.

If X follows a Poisson distribution with parameter λ (can be denoted as $X \sim \pi(\lambda)$), then

$$E[X] = Var[X] = \lambda$$

The m.g.f of Poisson distribution is $e^{\lambda(e^t - 1)}$

Approximation of Binomial Distribution

For a binomial distribution $B(n, p)$, when p is small($p \leq 0.1$) and n is large($n \geq 10$), it's PDF is approximately $\pi(np)$. That is to say,

$\binom{n}{x} p^x (1-p)^{n-x} \approx \frac{(np)^x e^{-np}}{x!}$

Outline

Additivity

Suppose X_1 and X_2 follows a certain distribution. If $X_1 + X_2$ also follow this kind of distribution, we say that this kind of distribution is additive. Almost every discrete probability distribution is additive.

1. Binomial distribution: If $X_1 \sim B(n_1, p), X_2 \sim B(n_2, p)$, then $(X_1 + X_2) \sim B(n_1 + n_2, p)$.

2. Pascal distribution: If X_1 follows a Pascal distribution with parameters p and r_1, X_2 follows a Pascal distribution with parameters p and r_2 ,then $X_1 + X_2$ follows a Pascal distribution with parameters p and $r_1 + r_2$.

3. Poisson distribution: If $X_1 \sim \pi(\lambda_1), X_2 \sim \pi(\lambda_2)$, then $(X_1 + X_2) \sim \pi(\lambda_1 + \lambda_2)$.

These properties can all be easily proved by calculating their product of m.g.f. The additivity also holds for the sum of n random variables. Notice that geometric distribution is Pascal distribution with r=1.

ex 2.3

1. Suppose X follows a geometric distribution with parameter 0.5. Take 10 i.i.d sample of X, what's the probability that their sum is smaller or equal to 15?

2. Prove the additivity of Poisson distribution **without** using its m.g.f.

Outline

Q&A

Please summarize all the discrete probability distribution we have learnt.
Write down their PDF, m.g.f as well as expectations and variance.