

MA2116 - Probability

Liew Zhao Wei

Semester 1, 2023-2024

Abstract

This is a set of incomplete notes for an introductory course on probability.

1 Preliminaries

1.1 Taylor series expansions

We should know the following Taylor series expansions by heart:

$$\begin{aligned}e^x &= \sum_{n=0}^{\infty} \frac{x^n}{n!} \\ \ln(1+x) &= \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n} \\ \sin x &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} \\ \cos x &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}\end{aligned}$$

2 Axioms of Probability

TODO.

3 Discrete random variables

3.1 Expectation

When given a random variable, we are often interested in its *average* value. This is captured by the notion of *expectation*.

Definition 3.1 (Expected value of a discrete r.v.)

Let X be a discrete random variable. Then the *expected value*, or *expectation*, of X , denoted by $E[X]$, is defined as

$$E[X] = \sum_{x \in \Omega_X} x \Pr(X = x).$$

Proposition 3.2 (Expectation of a function of a discrete r.v.)

Let X be a discrete random variable and g be real-valued a function. Then

$$E[g(X)] = \sum_{x \in \Omega_X} g(x) \Pr(X = x).$$

Corollary 3.3

Let X be a random variable and $a, b \in \mathbb{R}$. Then

$$\mathbb{E}[aX + b] = a \mathbb{E}[X] + b.$$

One important property of expectations of (general) random variables is that it is *linear*.

Theorem 3.4

Linearity of expectation Let X_1, X_2, \dots, X_n be random variables (not necessarily discrete). Then

$$\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i].$$

3.2 Variance

Expectation alone doesn't reveal much about the distribution of a random variable. For example, compare these two slot machines:

1. The first slot machine gives you \$100 with probability 1.
2. The second slot machine gives you \$10000 with probability 0.01 and \$0 with probability 0.99.

The expected payout of both slot machines is \$100, but the chances of winning are very different. This difference is captured by the notion of *variance*.

Definition 3.5 (Variance of a discrete r.v.)

Let X be a discrete random variable. Then the *variance* of X , denoted by $\text{Var}(X)$, is defined as

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Remark 3.6 (Alternative formula). *The variance can also be computed using the following formula:*

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Remark 3.7. *Since variance is non-negative, we have*

$$\mathbb{E}[X^2] \geq \mathbb{E}[X]^2.$$

This can be used to prove the friendship paradox.

Proposition 3.8 (Variance of linear combination)

Let X be a discrete random variable and $a, b \in \mathbb{R}$. Then

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

3.3 Bernoulli and binomial random variables

Perhaps the most basic and well-known random variable is the *binomial* random variable and its little brother, the *Bernoulli* random variable. We motivate this with the age old example of a coin toss.

Example 3.9 (Coin toss)

Suppose we toss a coin n times that comes up heads with probability p and tails with probability $1 - p$. Let X be the number of heads that appear. Then X is a *binomial* random variable with parameters n and p . It takes values in $\{0, 1, \dots, n\}$.

Definition 3.10 (Binomial random variable)

A random variable X is said to be *binomial* with parameters n and p , denoted by $X \sim \text{Bin}(n, p)$, if X has the following probability mass function:

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

Remark 3.11 (Bernoulli random variable). A Bernoulli random variable is a special case of a binomial random variable with $n = 1$.

Theorem 3.12 (Properties of binomial random variables)

Let $X \sim \text{Bin}(n, p)$. Then

1. $E[X^k] = np E[(Y + 1)^{k-1}]$ where $Y \sim \text{Bin}(n-1, p)$
2. $E[X] = np$
3. $\text{Var}(X) = np(1-p)$

Proof. The proof is by some algebraic manipulation of the definition of $E[X^k]$ and using the identity

$$i \binom{n}{i} = n \binom{n-1}{i-1}$$

□

The following relationship is useful for computing the distribution function of a binomial random variable. It can also be used to find the maximum of the probability mass function.

Proposition 3.13 (Relationship between $\Pr(X = k+1)$ and $\Pr(X = k)$)

Let $X \sim \text{Bin}(n, p)$. Then

$$\Pr(X = k+1) = \frac{p}{1-p} \frac{n-k}{k+1} \Pr(X = k).$$

Corollary 3.14 (Condition for monotonic increase)

Let $X \sim \text{Bin}(n, p)$. Then, as k goes from 0 to n , $\Pr(X = k)$ first increases monotonically and then decreases monotonically. In fact, $\Pr(X = k) \geq \Pr(X = k+1)$ if and only if $k \leq (n+1)p$.

3.4 Poisson random variable

Besides the binomial random variable, another important discrete random variable is the *Poisson* random variable.

Definition 3.15 (Poisson random variable)

A random variable X is said to be *Poisson* with parameter $\lambda > 0$, denoted by $X \sim \text{Pois}(\lambda)$, if X has the following probability mass function:

$$\Pr(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}, \quad i = 0, 1, 2, \dots$$

Remark 3.16. By the Taylor series expansion of e^x , we have

$$\sum_{i=0}^{\infty} \Pr(X = i) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1.$$

The Poisson random variable is useful for approximating the binomial random variable when n is large and p is small enough so that $\lambda = np$ is of moderate size.

Example 3.17 (Examples that can be approximated by a Poisson r.v.)

We can think of each of these examples as a long series of independent trial, each with a small chance of success. Thus, we can approximate the number of successes by a Poisson random variable.

1. The number of misprints on a page of a book.
2. The number of people in a community who survive to age 100.
3. The number of wrong telephone numbers dialed in a day.
4. The number of packages of biscuits sold in a store in a day.
5. The customers centering a post office on a given day.

Remark 3.18. *This approximation works decently even with n as small as 10 and p as large as 0.1. For example, with $X \sim \text{Bin}(10, 0.1)$ and $Y \sim \text{Pois}(1)$, we have*

$$\Pr(X \leq 1) \approx 0.7361$$

$$\Pr(Y \leq 1) \approx 0.7358$$

Theorem 3.19 (Properties of Poisson r.v.)

Let $X \sim \text{Pois}(\lambda)$. Then

1. $E[X] = \lambda$
2. $\text{Var}(X) = \lambda$

We have shown that the Poisson distribution approximates the binomial distribution when n is large and p is small. In fact, we do not need the events be independent or to occur with equal probability. We just need the probability of the events to be small and the events to be at most weakly dependent. This is explained in what we call the *Poisson paradigm*.

Definition 3.20 (Weak dependence)

Events E and F are weakly dependent if $\Pr(E) \approx \Pr(E | F)$.

Theorem 3.21 (Poisson Paradigm)

Consider n events where the i th event occurs with probability p_i . If all the p_i are small and the trials are either independent or at most weakly dependent, then the number of events that occur approximately has a Poisson distribution with mean $\lambda = \sum_{i=1}^n p_i$.

Proposition 3.22 (Relationship between $\Pr(X = k + 1)$ and $\Pr(X = k)$)

Let $X \sim \text{Pois}(\lambda)$. Then

$$\Pr(X = k + 1) = \frac{\lambda}{k + 1} \Pr(X = k).$$

One other use of the Poisson distribution is in *poisson processes*.

Theorem 3.23 (Poisson process)

Let $\lambda > 0$. Suppose events occur at random points in time where

1. The probability that exactly 1 event occurs in an interval of length h is approximately $\lambda h + o(h)$, where $o(h)$ stands for any function $f(h)$ such that $\lim_{h \rightarrow 0} f(h)/h = 0$.
2. The probability that more than 1 event occurs in an interval of length h is $o(h)$.
3. The number of events occurring in disjoint intervals are independent.

Then, the number of events occurring in an interval of length t , denoted by $N(t)$, is a Poisson random variable with parameter λt .

Example 3.24 (Examples of poisson processes)

Poisson processes can approximate phenomena like:

1. The number of earthquakes occurring in a fixed time span.
2. The number of wars per year.
3. The number of electrons emitted from a heated cathode during a fixed time period.
4. The number of deaths, in a given period of time, of the policyholders of a life insurance company.

3.5 Geometric random variable**3.6 Negative binomial random variable****3.7 Hypergeometric random variable****4 Continuous random variables**

TODO.

5 Jointly distributed random variables

TODO.

6 References

1. A First Course in Probability (10th edition) by Sheldon Ross