Chapter 3: General Probability and Random Variables

Wang Shujia

Contents

1 Random Events

1.1 Counting

Basic principle of counting

- Suppose k experiments are to be performed such that the first can result in any one of n_1 outcomes;
- and if for each of these n_1 outcomes, there are n_2 possible outcomes of the second experiment;
- and if for each of the possible outcomes of the first two experiments, there are n_3 possible outcomes of the third experiment;
- and if . . . ,

then there are

$$n = n_1 \times n_2 \times \cdots \times n_k$$

possible outcomes for the k experiments.

Permutation

Any ordered sequence of k objects taken from n distinct objects is called a **permutation** and is denoted $P_{k,n}$

$$P_{k,n} = n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

Example: How many different ways can the first 3 places be decided in a race with 5 runners?
Solution: in R,

```
> prod(5:(5-3+1))

## [1] 60

> factorial(5)/factorial(5-3)

## [1] 60
```

Combinations

An arrangement of k objects taken from n objects without regard to order is called a **combination** and is denoted $C_{k,n}$ or $\binom{n}{k}$.

$$C_{k,n} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Example: A committee of three people is to be formed from a group of eight people. How many different committees are possible?
Solution: in R,

```
> choose(n=8,k=3)

## [1] 56
```

1.2 Set Theory

Events

- A **random experiment** is any action or process that generates an outcome whose value cannot be known with certainty.
- The **sample space** of an experiment, denoted by Ω or S, is the set of all of the possible outcomes of an experiment.
- An **event** is any subset of the sample space, which is often denoted with the letter E.
- **Elementary Event**: Event that cannot be divided into subsets.

Basic Set Theory

Set is collection of elements. Let

$$A = \{a_1, a_2, \ldots, a_n\}, \ B = \{b \in \mathbb{R} : b > 1\}$$

- $\emptyset$: Empty set, has no elements
- Ω or S: Universal set (sample space), all elements we consider
- $A \subset B$: A is subset of B if every element of A is element of B
- Transitivity: $A \subset B$ and $B \subset C$ implies $A \subset C$
- Equality: $A = B$ if and only if $A \subset B$ and $B \subset A$
- Union:$A \cup B$ or $A + B$, set with elements belonging to A or B or both
- Intersection:$A \cap B$ or AB, set with elements belonging to A and to B
- A and B disjoint if :$A \cap B = \emptyset$, have no common elements
- Complement A^c or $\bar{A}$: all elements in Ω not in A

Basic Set Laws

1. Commutative laws
   - for the union: $A \cup B = B \cup A$
   - for the intersection: $A \cap B = B \cap A$
2. Associative laws
   - for the union: $(A \cup B) \cup C = A \cup (B \cup C)$
   - for the intersection: $(A \cap B) \cap C = A \cap (B \cap C)$
3. Distributive laws:
   - $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$
   - $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

4. DeMorgan's laws

$$(\bigcap_{i=1}^{\infty} A_i)^c = \bigcup_{i=1}^{\infty} A_i^c$$

$$(\bigcup_{i=1}^{\infty} A_i)^c = \bigcap_{i=1}^{\infty} A_i^c$$

Illustration with R

```
> S<-c(1,2,3,4,5,6); A<-c(1,2); B<-c(1,2,3); C<-c(2,4,6)
> intersect(A,B)

## [1] 1 2

> intersect(intersect(A,B),C) #works pairwise

## [1] 2

> union(A,B)

## [1] 1 2 3

> union(union(A,B),C) #works pairwise

## [1] 1 2 3 4 6

> (complementofA<-setdiff(S,A)) #complement of A

## [1] 3 4 5 6

> intersect(A,complementofA)

## numeric(0)
```

2 Probability

Ellsberg Paradox

- Two urns
  1. 100 balls with 50 red and 50 blue.
  2. A mix of red and blue but you don't know the proportion.
- Which urn would you like to bet on?
- People don't like the ”uncertainty” about the distribution of red/blue balls in the second urn.
- Probability is counter-intuitive!

2.1 Three Axioms of Probability

Definition of probability

Probability is any function $P(\cdot)$ of event $E \subset \Omega$ such that

1. $0 \leq P(E) \leq 1$ for any $E \subset \Omega$
2. $P(\Omega) = 1$
3. For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is $E_i \cap E_j = \emptyset$ for all $i \neq j$,

$$P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$$

Properties

1. $P(E^c) = 1 - P(E)$
2. $P(\emptyset) = 0$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. If $A \subset B$, then $P(A) \leq P(B)$

Birthday Problem

Suppose that a room contains m students. What is the probability that at least two of them have the same birthday?

Solution: Let the event E denote two or more students with the same birthday,

$$P(E) = 1 - P(E^c) = 1 - \frac{365 \times 364 \times \cdots \times (365 - m + 1)}{365^m}$$

```
> m<-seq(10,50,5)
> P.E<-function(m){
+      c(Students=m,ProbAtL2SB=1-prod((365:(365-m+1)/365)))}
> round(sapply(m,P.E),2)

##             [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9]
## Students   10.00 15.00 20.00 25.00 30.00 35.00 40.00 45.00 50.00
## ProbAtL2SB  0.12  0.25  0.41  0.57  0.71  0.81  0.89  0.94  0.97
```

2.2 Conditional Probability and Independent Events

Definition of Conditional Probability

If A and E are any two events in a sample space Ω and $P(E) > 0$, the **conditional probability** of A given E is defined as

$$P(A|E) = \frac{P(A \cap E)}{P(E)}$$

Interpretation:

$$P(A|E) = P(A \text{ holds } \textbf{\textit{given}} \text{ that } B \text{ holds})$$

- Multiplication Rule: for $P(A) > 0, P(B) > 0$,

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

Independent Events

- Two events E and F are **independent** if and only if
  $P(E|F) = P(E)$ and $P(F|E) = P(F)$.
- Equivalently, two events E and F are **independent** if and only if
  $P(E \cap F) = P(E)P(F)$
- Events $E_1, E_2, \ldots, E_n$ are **independent** if, for *every* k where $k = 2, \ldots, n$ and every subset of indices $i_1, i_2, \ldots, i_k$,

$$P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{ik}) = P(E_{i_1})P(E_{i_2}) \cdots P(E_{i_k})$$

- How many equations there exist?

Realistic Example

D disease, $\bar{D}$ not have disease; T_+ test positive, T_- test negative

		Test Positive T_+	Test Negative T_-	Total
disease	D	.008	.002	.01
healthy	$\bar{D}$	.0063	.985	.99
Total		.0143	.9857	1.00

$$P(D \cap T_+) = .008$$

$$P(D|T_+) = \frac{P(D \cap T_+)}{P(T_+)} = \frac{.008}{.0143} = .559$$

2.3 The Law of Total Probability and Bayes' Rule

Law of Total Probability

Theorem

Let $F_1, F_2, \ldots, F_n$ be such that $\bigcup_{i=1}^n F_i = \Omega$ and $F_i \cap F_j = \emptyset$ for all $i \neq j$, with $P(F_i) > 0$ for all i. Then, for any event A,

$$P(A) = \sum_{i=1}^{n} P(F_i \cap A) = \sum_{i=1}^{n} P(F_i)P(A|F_i)$$

- $P(A) = P(E)P(A|E) + P(\bar{E})P(A|\bar{E})$

Bayes's Rule

Theorem

Let $F_1, F_2, \ldots, F_n$ be such that $\bigcup_{i=1}^n F_i = \Omega$ and $F_i \cap F_j = \emptyset$ for all $i \neq j$, with $P(F_i) > 0$ for all i. Then, for any event A,

$$P(F_j|A) = \frac{P(A \cap F_j)}{P(A)} = \frac{P(F_j)P(A|F_j)}{\sum_{i=1}^n P(F_i)P(A|F_i)} \text{ for } j = 1, 2, \ldots, n$$

Famous example: Monty Hall problem

Problem is named after the host of the long-running TV show, *Let's make a Deal.*

- *Marilyn vos Savant* had a column with the correct answer that many Mathematicians thought was wrong!
- A *contestant* is given the choice of 3 doors.
- There is a *prize (a new car, say)* behind one of the doors and something worthless behind the other two doors: two goats.

You pick a door. What is your probability of winning?

Let$A =$"door you picked has the car behind it", then $P(A) = 1/3$.

Puzzle

Now the twist.

The host - who knows what's behind the doors - opens one of the other two doors to reveal a goat.

He asks "would you like to switch doors?"

Which strategy do you chose?

Solution 1: Simulation

- **actual**: Generated random vector of size 10,000 containing the numbers 1, 2, and 3 to represent the door behind which the car is contained.
- **guess**: Generated random vector of size 10,000 containing the numbers 1, 2, and 3 to represent the contestant's initial guess.

```
> set.seed(2) # done for reproducibility
> actual <- sample(x = 1:3, size = 10000, replace = TRUE)
> aguess <- sample(x = 1:3, size = 10000, replace = TRUE)
> equals <- (actual == aguess)
> PNoSwitch <- sum(equals)/10000
> not.eq <- (actual != aguess)
> PSwitch <- sum(not.eq)/10000
> Probs <- c(PNoSwitch, PSwitch)
> names(Probs) <- c("P(Win no Switch)", "P(Win Switch)")
> Probs

## P(Win no Switch)    P(Win Switch)
##           0.3348           0.6652
```

Solution 2: Total Probability

Let $B =$"Win the car if switch",

then from the Law of Total Probability

$$\begin{aligned} P(B) &= P(A)P(B|A) + P(\bar{A})P(B|\bar{A}) \\ &= \frac{1}{3} \times 0 + \frac{2}{3} \times 1 = \frac{2}{3} \end{aligned}$$

Solution 3: Bayes Theorem

Assume you initially guesses Door 1 and that Monty opens Door 3.
Let the event D_i =Door i conceals the prize ($i = 1, 2, 3$) and
O_j =Monty opens door j after the contestant selects Door 1 ($j = 3$ here).
Compare $P(D_1|O_3)$ and $P(D_2|O_3)$:

$$\begin{aligned}
P(D_1|O_3) &= \frac{P(D_1)P(O_3|D_1)}{\sum_{j=1}^{3} P(D_j)P(O_3|D_j)} \\
&= \frac{1/2 \times 1/3}{1/2 \times 1/3 + 1 \times 1/3 + 0 \times 1/3} = \frac{1}{3} \\
P(D_2|O_3) &= \frac{P(D_2)P(O_3|D_2)}{\sum_{j=1}^{3} P(D_j)P(O_3|D_j)} \\
&= \frac{1 \times 1/3}{1/2 \times 1/3 + 1 \times 1/3 + 0 \times 1/3} = \frac{2}{3}
\end{aligned}$$

3 Random Variables

Definition of Random Variables

A **random variable** is a function from a sample space Ω into the real numbers.

- **Variable** has a numerical value is determined by the outcome of a random experiment. *A number that hasn't happened yet*
- **Data**: Once random variables "happen" they then become data. Then you can learn about future random variables.
- **Probability Distribution**: This is a function that tells you how likely it is that each value will occur
- **Expected value** and **Variance** summarize probability distributions just like sample means and variances do for data.

3.1 Discrete Random Variables

Probability Density Functions

Discrete random variables are characterized by their *supports* which take the form $S = \{x_1, x_2, x_3, \ldots, \}$

Definition: The function that assigns probability to the values of the random variable is called the *probability density function* (pdf) or *probability mass function* (pmf)

$$f(x) = P(X = x), \quad x \in S$$

All pdfs must satisfy:

1. $f(x) \geq 0$ for all x
2. $\sum f(x) = 1$

Example

Toss a coin 3 times.
The sample space would be

$$\Omega = \{HHH, HTH, THH, TTH, HHT, HTT, THT, TTT\}$$

Now let X be the number of Heads observed.
Then X has support $S = \{0, 1, 2, 3\}$.
We can represent the pmf with a table:

x	0	1	2	3	Total
$f(x)$	1/8	3/8	3/8	1/8	1

Cumulative Distribution Function

The **cumulative distribution function** (cdf) is defined as

$$F(x) = P(X \leq x) = \sum_{k \leq x} P(X = k), \ -\infty < x < \infty$$

Properties:

1. $0 \leq F(x) \leq 1$
2. If $a < b$, then $F(a) \leq F(b)$ for any real numbers a and b. In other words, $F(x)$ is a non-decreasing function of x
3. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$
4. $F(x)$ is a step function, and the height of the step at x is equal to $f(x) = P(X = x)$

Mean, Variance, and Standard Deviation

Mean (Expectation) $\mu = \mathrm{E}(X) = \sum_{x \in S} x f_X(x)$

Variance $\sigma^2 = \mathrm{E}(X - \mu)^2 = \sum_{x \in S} (x - \mu)^2 f_X(x)$

Standard Deviation $\sigma = \sqrt{\sigma^2}$

Usefull Formula $\sigma^2 = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$

```
> x<-c(0,1,2,3)
> f<-c(1/8,3/8,3/8,1/8)
> mu<-sum(x*f)
> sigma2<-sum((x-mu)^2*f)
> sigma<-sqrt(sigma2)
> mu;sigma2;sigma

## [1] 1.5
## [1] 0.75
## [1] 0.8660254
```

The mean μ =1.5; variance σ^2 =0.75

3.2 Continuous Distributions

Probability Density Functions

Every continuous random variable X has a **probability density function** (pdf) that satisfies three basic properties:

1. $f(x) \geq 0, \ -\infty < x < \infty$
2. $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$
3. $P(X \in A) = \int_{x \in A} f(x)\mathrm{d}x, \ \text{for an event } A \subset \Omega$

The **support** of $f(x)$: $S = \{x | f(x) > 0\}$.

We say that $X \sim f(x)$.

Cumulative Distribution Function

The **cumulative distribution function** (cdf) is defined by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$$

Properties:

1. $F(x)$ is nondecreasing
2. $F(x)$ is continuous, and $F^{'}(x) = f(x)$
3. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$

Expectation and Variance

For a continuous random variable X the **expected value** of $g(X)$ is

$$\mathrm{E}[g(X)] = \int g(x)f(x)\mathrm{d}x$$

The expectation of random variable X

$$\mu = \mathrm{E}(X) = \int xf(x)\mathrm{d}x$$

The variance of X

$$\sigma^2 = \mathrm{E}(X - \mu)^2 = \int (x - \mu)^2 f(x)\mathrm{d}x$$

Numerical Integration with R

The R function integrate() approximates the *integral of functions* of one variable over a finite or infinite interval and estimates the absolute error in the approximation.

Example: compute

$$P(-0.5 \leq X \leq 1) = \int_{-0.5}^{1} \frac{3}{4}(1 - x^2)\mathrm{d}x = 0.84375$$

```
> fx <- function(x) {
+       3/4 - 3/4 * x^2 } # define function fx
> integrate(fx,lower=-0.5,upper=1) # gives value and tolerance

## 0.84375 with absolute error < 9.4e-15
```

3.3 Moment Generating Functions

Moment Generating Functions

The **moment generating function** (mgf) of a r.v.X is defined by

$$M_X(t) = \mathrm{E}(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\mathrm{d}x$$

for any t such that the integral exists.

If X is discrete, then

$$M_X(t) = \sum_x e^{tx} p(x)$$

Example: mgf of Binomial

Find the MGF for $X \sim \mathrm{Bin}(n,p)$.

$$\begin{aligned} M(t) &= \sum_{\mathrm{i=0}}^{\mathrm{n}} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{\mathrm{i=0}}^{\mathrm{n}} \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= (pe^t + q)^n \end{aligned}$$

Example: mgf of Normal

If $X \sim \mathrm{N}(\mu, \sigma^2)$, then the mgf is

$$\begin{aligned} M(t) &= \mathrm{E}(e^{tX}) \\ &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x \\ &= e^{\mu t + \frac{\sigma^2}{2} t^2} \end{aligned}$$

Application 1: Determine the distribution of a r.v.

Theorem 1. *The moment generating function, if it exists in a neighborhood of zero, determines a probability distribution uniquely.*

- The proof is beyond the scope of our class.
- An MGF is also known as a "Laplace Transform" and is manipulated in that context in many branches of science and engineering.
- This is the key tool to proof the well known theorem of central limit.

Application 2: Compute High Order Moments

Why is it called a Moment Generating Function?

$$\begin{aligned} M^{'}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\sum_{x\in S} e^{tx}f(x) = \sum_{x\in S} e^{tx}xf(x) \\ M^{'}(0) &= \sum_{x\in S} e^{0}xf(x) = \mathrm{E}(X) \\ M^{''}(0) &= \mathrm{E}(X^2) \\ &\cdots \\ M^{(k)}(0) &= \mathrm{E}(X^k) \end{aligned}$$

Example: Mean and Variance of Binomial

Find the mean and variance of `binom(`n`,`p`)`.

$$\begin{aligned} M(t) &= (pe^t + q)^n \\ M^{'}(0) &= n(pe^t + q)^{n-1}pe^t|_{t=0} = np \\ M^{''}(0) &= [n(n-1)(pe^t + q)^{n-2}(pe^t)^2 + n(pe^t + q)^{n-1}pe^t]|_{t=0} \\ &= n(n-1)p^2 + np \end{aligned}$$

The expectation $\mathrm{E}(X) = np$ and the variance

$$\mathrm{D}(X) = \mathrm{E}(X^2) - (\mathrm{E}X)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

Expansion of a MGF

Taylor series expansion:

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k, \text{ for all } |x-a| < R$$

where R is called the *radius of convergence* of the series.

Theorem 2. *If an MGF exists for all t in the interval $(-\epsilon, \epsilon)$, then*

$$M(t) = \sum_{k=0}^{\infty} \frac{\mathrm{E}(X^k)}{k!}t^k, \text{ for all } |t| < \epsilon$$

3.4 Transformations of Random Variables

Functions of Continuous Random Variables

The *goal* here is to determine the distribution of $U = g(X)$ based on the distribution of X.

General approaches:

1. pdf transformation
2. calculate the cdf
3. by mgf

The pdf transformation

Theorem 3. *Let X have pdf $f_X(x)$ and let $U = g(X)$ be a function of X which is one-to-one with a differentiable inverse g^{-1}. Then the pdf of $U = g(X)$ is given by*

$$f_U(u) = f_X(x)\left|\frac{\mathrm{d}x}{\mathrm{d}u}\right|$$

where $x = g^{-1}(u)$ is the inverse function of g.

Example: pdf of Logmormal

Let $X \sim N(\mu, \sigma^2)$, and let $Y = e^X$. What is the pdf of Y?

Solution: Notice first that $e^x > 0$ for any x, so the support of Y is $(0, \infty)$. Since the

transformation is monotone, we can solve $y = e^x$ for x to get $x = \ln y$, giving $dx/dy = 1/y$. Therefore, for any $y > 0$,

$$\begin{aligned} f_Y(y) &= f_X(x)\left|\frac{1}{y}\right| \\ &= \frac{1}{y\sigma\sqrt{2\pi}}\exp\left\{\frac{(\ln y - \mu)^2}{2\sigma^2}\right\} \end{aligned}$$

The pdf implies $Y \sim \text{lnorm}(\mu, \sigma^2)$

The cdf method

Example 4. Suppose $X \sim U(0, 1)$ and suppose that we let $Y = -\ln X$. What is the PDF of Y ?

Solution: The support set of X is $(0, 1)$, and y traverses $(0, \infty)$ as x ranges from 0 to 1, so the support set of Y is $S_Y = (0, \infty)$. For any $y > 0$, we consider

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(-\ln X \le y) = P(X \ge e^{-y}) \\ &= 1 - F_X(e^{-y}) \ (0 < e^{-y} < 1 \text{ for } y > 0) \\ &= 1 - e^{-y} \end{aligned}$$

The pdf of Y is $f_Y(y) = F_Y^{'}(y) = e^{-y}$, for $y > 0$. This implies $Y \sim \exp(1)$.

The mgf method

Example 5. Suppose $X \sim N(0, 1)$, $Y \sim N(0, 1)$, what is the distribution of $X + Y$ if X and Y are independent?

Solution: The mgf of $N(\mu, \sigma^2)$:

$$m(t) = e^{\mu t + \frac{\sigma^2}{2}t^2}$$

The mgf of $Z = X + Y$

$$\begin{aligned} m_Z(t) &= \mathrm{E}(e^{t(X+Y)}) = m_X(t)m_Y(t) \\ &= e^{\frac{t^2}{2}} \cdot e^{\frac{t^2}{2}} = e^{\frac{2t^2}{2}} \end{aligned}$$

This mgf implies $Z \sim N(0, 2)$.