36225
Probability

ZhenTing Liu

Contents

Chapter 1

Multivariate Distribution

1.1 Multivariate distribution

Definition 1.1.1

Let Y_1 and Y_2 be discrete random variables. The joint (or bivariate) probability function for Y_1 and Y_2 is given by

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2), \quad -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

In the single-variable case , the probability function for a discrete random variable Y assigns nonzero probabilities to a finite or countable number of distinct values of Y in such a way that the sum of the probabilities is equal to 1. Similarly, in the bivariate case the joint probability function $p(y_1, y_2)$ assigns nonzero probabilities to only a finite or countable number of pairs of values (y_1, y_2). Further, the nonzero probabilities must sum to 1.

Theorem 1.1.1 Discrete random variables

If Y_1 and Y_2 are discrete random variables with joint probability function $p(y_1, y_2)$, then

1. $p(y_1, y_2) \geqslant 0 \quad \forall y_1, y_2$
2. $\sum_{y_1, y_2} p(y_1, y_2) = 1$

As in the case of univariate random variables, the distinction between jointly discrete and jointly continuous random variables may be characterized in terms of their (joint) distribution functions.

Definition 1.1.2

For any random variables Y_1 and Y_2, the joint (bivariate) distribution function $F(y_1, y_2)$ is

$$F(y_1, y_2) = P(Y_1 \leqslant y_1, Y_2 \leqslant y_2) \quad -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

Note:-

If Y_1 and Y_2 are random variables with joint distribution function $F(y_1, y_2)$, then

1. $F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0$
2. $F(\infty, \infty) = 1$

Theorem 1.1.2 Continous random variables

If Y_1 and Y_2 are jointly continuous random variables with a joint density function given by $f(y_1, y_2)$, then

1. $f(y_1, y_2) \geqslant 0, \quad$ for all $\quad y_1, y_2$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2 = 1$

1.2 Marginal and Conditional distribution

univariate event $(Y_1 = y_1)$ is the union of bivariate events of the type $(Y_1 = y_1, Y_2 = y_2)$, with the union being taken over all possible values for y_2.

Definition 1.2.1: Marginal Distribution

Let Y_1 and Y_2 be jointly discrete random variables with probability function $p(y_1, y_2)$. Then the marginal probability functions of Y_1 and Y_2, respectively, are given by

$$p_1(y_1) = \sum_{\text{all} y_2} p(y_1, y_2) \quad \text{and} \quad p_2(y_2) = \sum_{\text{all} y_1} p(y_1, y_2)$$

Let Y_1 and Y_2 be jointly continuous random variables with joint density function $f(y_1, y_2)$. Then the marginal density functions of Y_1 and Y_2, respectively, are given by

$$f_1(y_1) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_2 \quad \text{and} \quad f_2(y_2) = \int_{-\infty}^{\infty} f(y_1, y_2) dy_1$$

The term **marginal**, as applied to the univariate probability functions of Y_1 and Y_2 , has intuitive meaning. To find $p_1(y_1)$, we sum $p(y_1, y_2)$ over all values of y_2, and hence accumulate the probabilities on the y_1 axis (or margin).

Note:-

To find marginal pdf

1. integrate over the other variable
2. fix a line corresponding to a value of desired marginal pdf
3. find the limit of integral: intersecting points between this line and support of the joint pdf
4. check if the boundary of support is straight line
5. remember to specify the support of the resulting marginal pdf

Definition 1.2.2: Conditional Distribution discrete case

If Y_1 and Y_2 are jointly discrete random variables with joint probability function $p(y_1, y_2)$ and marginal probability functions $p_1(y_1)$ and $p_2(y_2)$, respectively, then the conditional discrete probability function of Y_1 given Y_2 is

$$p(y_1|y_2) = P(Y_1 = y_1|Y_2 = y_2) = \frac{P(Y_1 = y_1, Y_2 = y_2)}{P(Y_2 = y_2)} = \frac{p(y_1, y_2)}{p_2(y_2)}$$

Definition 1.2.3: Conditional Distribution cont. case

If Y_1 and Y_2 are jointly continuous random variables with joint density function $f(y_1, y_2)$ and marginal densities $f_1(y_1)$ and $f_2(y_2)$ respectively, then the conditional distribution function of Y_1 given $Y_2 = y_2$ is

$$\begin{aligned} F(y_1|y_2) &= P(Y_1 \leqslant y_1|Y_2 = y_2) \\ &= \int_{-\infty}^{y_1} \frac{f(t_1, y_2)}{f_2(y_2)} dt_1 \end{aligned}$$

We will call the integrand of this expression the conditional density function of Y_1 given $Y_2 = y_2$ and we will denote it by $f(y_1|y_2)$.
Note for any $f_2(y_2) > 0$, the conditional density of Y_1 given $Y_2 = y_2$ is given by

$$f(y_1|y_2) = \frac{f(y_1, y_2)}{f_2(y_2)}$$

and, for any $f_1(y_1) > 0$, the conditional density of Y_2 given $Y_1 = y_1$ is gievn by

$$f(y_2|y_1) = \frac{f(y_1, y_2)}{f_1(y_1)}$$

Notice that $F(y_1|y_2)$ is a function of y_1 for a fixed value of y_2

Note:-

$$P(a < Y_1 < b|Y_2 = y_2) = \int_a^b f(y_1|y_2)dy_1$$

also for a valid pmf/pdf for Y_1 or Y_2 as univariate r.v.

1. $f(y_1|y_2) = \frac{f(y_1,y_2)}{f_2(y_2)} \geqslant 0$
2. $\int_{-\infty}^{\infty} f(y_1|y_2)dy_1 = \int_{-\infty}^{\infty} \frac{f(y_1,y_2)}{f_2(y_2)} dy_1 = 1$

1.2.1 Examples

$$f(y_1, y_2) = \begin{cases} 30y_1y_2^2, & y_1 - 1 \leqslant y_2 \leqslant 1 - y_1,\ 0 \leqslant y_1 \leqslant 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Question 1:

a show that the marginal density of Y_1 is a beta density with $\alpha = 2$ and $\beta = 4$
b Derive the marginal density of Y_2
c Derive the conditional density of Y_2 given $Y_1 = y_1$
d Find $P(Y_2 > 0|Y_1 = 0.75)$

Solution:
a

$$f_1(y_1) = \begin{cases} \int_{y_1-1}^{1-y_1} 30y_1y_2^2 dy_2, & 0 \leqslant y_1 \leqslant 1 \\ 0, & \text{elsewhere.} \end{cases}$$

~ Beta(2,4) after integration
b

$$f_2(y_2) = \begin{cases} \int_0^{1-y_2} 30y_1y_2^2 dy_1, & 0 \leqslant y_2 \leqslant 1 \\ \int_0^{y_2+1} 30y_1y_2^2 dy_1, & -1 \leqslant y_2 \leqslant 0 \\ 0, & \text{elsewhere.} \end{cases}$$

c

$$f(y_2|y_1) = \frac{f(y_1,y_1)}{f_1(y_1)} = \frac{30y_1y_2^2}{20y_1(1-y_1)^3} = \frac{3y_2^2}{2(1-y_1)^3} \quad y_1 - 1 \leqslant y_2 \leqslant 1 - y_1, 0 \leqslant y_1 \leqslant 1$$

d

$$\int_0^{(1-0.75)} \frac{3y_2^2}{2(1-0.75)^3} dy_2 = \frac{1}{2}$$

1.3 Independence

Definition 1.3.1: The r.v. Y_1, Y_2 are independent if and only if

discrete:

$$p(y_1,y_2) = p_1(y_1)p_2(y_2) \quad \forall y_1, y_2 \in \mathbb{R}^2$$

coninuous:

$$f(y_1,y_2) = f_1(y_1)f_2(y_2) \quad \forall y_1, y_2 \in \mathbb{R}^2$$

CDF:

$$F(y_1,y_2) = F_1(y_1)F_2(y_2) \quad \forall y_1, y_2 \in \mathbb{R}^2$$

Theorem 1.3.1 Determining independence

Given two r.v.'s Y_1, Y_2 Step 1: is the support of joint pmf/pdf rectangular with sides parallel to the axis?

- "No" $\implies$ Not independent
- "Yes" $\implies$ Move to step 2

Step 2: Can the joint pmf/pdf be written as the product of the form $p(y_1,y_2) = h(y_a)g(y_b), a, b \in \{1, 2\}$

- "No" $\implies$ Not independent
- "Yes" $\implies$ Independent

tip: To rigorously prove "cannot be factorized", one way is to find one pair of $(y_1, y_2) \in \mathbb{R}^2$ such that

$$f_1(y_1)f_2(y_2) \neq f(y_1,y_2)$$

1.4 Expected Value

Definition 1.4.1

Let $g(Y_1, Y_2, \dots, Y_k)$ be a function of the discrete random variables, $Y_1, Y_2, \dots, Y_k$, which have probability function $p(y_1, y_2, \dots, y_k)$. Then the *expected value* of $g(Y_1, Y_2, \dots, Y_k)$ is

$$E[g(Y_1, Y_2, \dots, Y_k)] = \sum_{\text{all } y_k} \cdots \sum_{\text{all } y_2} \sum_{\text{all } y_1} g(y_1, y_2, \dots, y_k)p(y_1, y_2, \dots, y_k).$$

If $Y_1, Y_2, \dots, Y_k$ are continuous random variables with joint density function $f(y_1, y_2, \dots, y_k)$, then

$$E[g(Y_1, Y_2, \dots, Y_k)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(y_1, y_2, \dots, y_k) f(y_1, y_2, \dots, y_k)\, dy_1\, dy_2 \cdots dy_k.$$

Note:-

In general, $E(Y_1Y_2) \neq E(Y_1)E(Y_2)$ except for when Y_1 and Y_2 are independent
Important: Independence $\Longrightarrow E(Y_1Y_2) = E(Y_1)E(Y_2)$ but $E(Y_1Y_2) = E(Y_1)E(Y_2) \not\Longrightarrow$ Independence

Consider two random variables Y_1, Y_2 with density function $f(y_1, y_2)$.

$$E(Y_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 f(y_1, y_2) dy_2 dy_1$$

1.5 Covariance

The covariance between two random variables Y_1 and Y_2, a measure of dependence between Y_1, Y_2, is defined as

$$cov(Y_1, Y_2) = E[(Y_1 - \mu_1)(Y_2 - \mu_2)] \quad \mu_i = E(Y_i) \quad i = 1, 2$$

which can also be written in shortcut form as

$$cov(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2)$$

note if $Y_1 = Y_2 = Y$ then $cov(Y, Y) = V(Y)$

Corollary 1.5.1

if Y_1, Y_2 are independent, then $cov(Y_1, Y_2) = 0$ because

$$E(Y_1Y_2) = E(Y_1)E(Y_2)$$

but $cov(Y_1Y_2) = 0$ does not imply independence

The larger the absolute value of the covariance of Y_1 and Y_2 , the greater the linear dependence between Y_1 and Y_2 . Positive values indicate that Y_1 increases as Y_2 increases; negative values indicate that Y_1 decreases as Y_2 increases. A zero value of the covariance indicates that the variables are uncorrelated and that there is no linear dependence between Y_1 and Y_2

1.5.1 Correlation Coefficient

correlation coefficient ρ is a measurement of dependency and is defined as

$$\rho = \frac{Cov(Y_1Y_2)}{\sigma_1\sigma_2}$$

where σ_1, σ_2 are the standard deviation of Y_1, Y_2 respectively.
Note

$$-1 \leqslant \rho_{Y_1Y_2} \leqslant 1$$

- $\rho = 0 \iff Y_1, Y_2$ uncorrelated
- $\rho = 1 \iff$ Perfect Positive linear dependence $Y_1 = aY_2 + b, \quad a > 0$
- $\rho = -1 \iff$ Perfect Negative linear dependence $Y_1 = -aY_2 + b, \quad a > 0$

Chapter 2

2.1 Eigentheory

Let's begin by talking about why we care about this topic

Consider the following System

Question 2

A particle jumps between A and B, suppose this particle starts at A, what are the probability that it ends up at A after

1. one step
2. n steps
3. ∞ steps

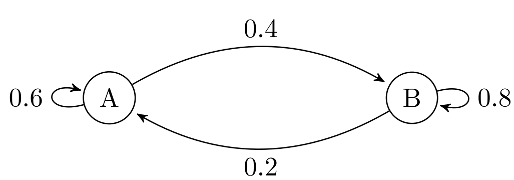
Solution: let $M = \begin{array}{c c} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 0.6 & 0.2 \\ 0.4 & 0.8 \end{pmatrix} \end{array}$ and $P_0 = \begin{array}{c c} \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 1 \\ 0 \end{pmatrix} \end{array}$ where M is called a stochastic Matrix, row 1 of P_0 is 1 because the particle is initially at A. So to answer (1), we simply multiply M and P_0, let $P_1 = MP_0$, we have $P_1 = \begin{array}{c c} \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix} \end{array}$

For (2), notice that we just multiply all the "states" together

$$\begin{aligned} P_n &= MP_{n-1} && (2.1) \\ &= M(MP_{n-2}) && (2.2) \\ &= \underbrace{M \dots M}_{n \text{ times}} P_0 = M^n P_0 && (2.3) \end{aligned}$$

The answer to (3) is a bit tricky, which we will discuss after seeing some tools to simplify our computation

Claim 2.1.1 $\exists$ invertible matrix X and diagnol matrix Λ such that $M = X\Lambda X^{-1}$



$$
\begin{aligned}
M &= X\Lambda X^{-1} \\
M^2 &= M(M) \\
&= (X\Lambda X^{-1})(X\Lambda X^{-1}) \\
&= X\Lambda^2 X^{-1}
\end{aligned}
$$

hence we must have

$$M^n = X\Lambda^n X^{-1}$$

This result is very useful to us because we essentially avoided calculating very large power of M, instead, we take the diagonal matrix Λ to its nth power

Note:-

Let $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ Recall that exponent of a diagonal matrix is just its entries to the same power, so

$$\Lambda^n = \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix}$$

Now, going back to (3), the computation is skipped here as we haven't talked must about it yet but we will discuss more later on. Λ for the problem happens to be $\begin{pmatrix} 1 & 0 \\ 0 & 0.4 \end{pmatrix}$ and X is $\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}$

$$\lim_{n\to\infty} M^n P_0 = \lim_{n\to\infty} \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 1^n & 0 \\ 0 & 0.4^n \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.67 \end{pmatrix}$$

So the probablity of the particle being at A after infinitely many steps is roughly 0.5

Definition 2.1.1: Limit of Sequence in $\mathbb{R}$

Let $\{s_n\}$ be a sequence in $\mathbb{R}$. We say

$$\lim_{n\to\infty} s_n = s$$

where $s \in \mathbb{R}$ if $\forall$ real numbers $\epsilon > 0$ $\exists$ natural number N such that for $n > N$

$$s - \epsilon < s_n < s + \epsilon \text{ i.e. } |s - s_n| < \epsilon$$

Question 3

Is the set x−axis$\setminus\{$Origin$\}$ a closed set

Solution: We have to take its complement and check whether that set is a open set i.e. if it is a union of open balls

Note:-

We will do topology in Normed Linear Space (Mainly $\mathbb{R}^n$ and occasionally $\mathbb{C}^n$)using the language of Metric Space

Claim 2.1.2 Topology

Topology is cool

Example 2.1.1 (Open Set and Close Set)

Open Set:
- ϕ
- $\bigcup\limits_{x \in X} B_r(x)$ (Any $r > 0$ will do)
- $B_r(x)$ is open

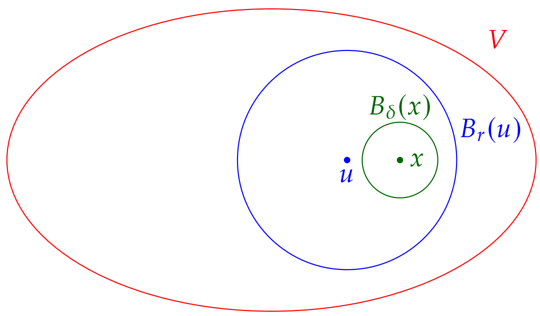
Closed Set:
- $X,\ \phi$
- $\overline{B_r(x)}$

$x-$axis $\cup$ $y-$axis

Theorem 2.1.1

If $x \in$ open set V then $\exists\ \delta > 0$ such that $B_\delta(x) \subset V$

Proof: By openness of V, $x \in B_r(u) \subset V$



Given $x \in B_r(u) \subset V$, we want $\delta > 0$ such that $x \in B_\delta(x) \subset B_r(u) \subset V$. Let $d = d(u,x)$. Choose δ such that $d + \delta < r$ (e.g. $\delta < \frac{r-d}{2}$)

If $y \in B_\delta(x)$ we will be done by showing that $d(u,y) < r$ but

$$d(u,y) \leqslant d(u,x) + d(x,y) < d + \delta < r$$

☺

Corollary 2.1.1

By the result of the proof, we can then show...

Lenma 2.1.1

Suppose $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{R}^n$ is subspace of $\mathbb{R}^n$.

Proposition 2.1.1

$1 + 1 = 2$.

2.2 Random

Definition 2.2.1: Normed Linear Space and Norm $\|\cdot\|$

Let V be a vector space over $\mathbb{R}$ (or $\mathbb{C}$). A norm on V is function $\|\cdot\|$ $V \longrightarrow \mathbb{R}_{\geqslant 0}$ satisfying

① $\|x\| = 0 \iff x = 0 \;\forall\; x \in V$

② $\|\lambda x\| = |\lambda|\|x\| \;\forall\; \lambda \in \mathbb{R}(\text{or } \mathbb{C}),\; x \in V$

③ $\|x+y\| \leqslant \|x\| + \|y\| \;\forall\; x, y \in V$ (Triangle Inequality/Subadditivity)

And V is called a normed linear space.
• Same definition works with V a vector space over $\mathbb{C}$ (again $\|\cdot\| \longrightarrow \mathbb{R}_{\geqslant 0}$) where ② becomes $\|\lambda x\| = |\lambda|\|x\|$ $\forall\; \lambda \in \mathbb{C},\; x \in V$, where for $\lambda = a + ib$, $|\lambda| = \sqrt{a^2+b^2}$

Example 2.2.1 ($p-$Norm)
$V = \mathbb{R}^m$, $p \in \mathbb{R}_{\geqslant 0}$. Define for $x = (x_1, x_2, \cdots, x_m) \in \mathbb{R}^m$

$$\|x\|_p = \left(|x_1|^p + |x_2|^p + \cdots + |x_m|^p\right)^{\frac{1}{p}}$$

(In school $p = 2$)

Special Case $p = 1$: $\|x\|_1 = |x_1| + |x_2| + \cdots + |x_m|$ is clearly a norm by usual triangle inequality.
Special Case $p \to \infty$ ($\mathbb{R}^m$ with $\|\cdot\|_\infty$): $\|x\|_\infty = \max\{|x_1|, |x_2|, \cdots, |x_m|\}$
For $m = 1$ these $p-$norms are nothing but $|x|$. Now exercise

Question 4

Prove that triangle inequality is true if $p \geqslant 1$ for $p-$norms. (What goes wrong for $p < 1$?)

Solution: **For Property ③ for norm-2**

When field is $\mathbb{R}$:

We have to show

$$\sum_i (x_i + y_i)^2 \leqslant \left(\sqrt{\sum_i x_i^2} + \sqrt{\sum_i y_i^2}\right)^2$$

$$\implies \sum_i (x_i^2 + 2x_iy_i + y_i^2) \leqslant \sum_i x_i^2 + 2\sqrt{\left[\sum_i x_i^2\right]\left[\sum_i y_i^2\right]} + \sum_i y_i^2$$

$$\implies \left[\sum_i x_iy_i\right]^2 \leqslant \left[\sum_i x_i^2\right]\left[\sum_i y_i^2\right]$$

So in other words prove $\langle x, y\rangle^2 \leqslant \langle x, x\rangle\langle y, y\rangle$ where

$$\langle x, y\rangle = \sum_i x_iy_i$$

Note:-

- $\|x\|^2 = \langle x, x\rangle$
- $\langle x, y\rangle = \langle y, x\rangle$

- $\langle \cdot, \cdot \rangle$ is $\mathbb{R}-$linear in each slot i.e.

$$\langle rx + x', y \rangle = r\langle x, y \rangle + \langle x', y \rangle \text{ and similarly for second slot}$$

Here in $\langle x, y \rangle$ x is in first slot and y is in second slot.

Now the statement is just the Cauchy-Schwartz Inequality. For proof

$$\langle x, y \rangle^2 \leqslant \langle x, x \rangle \langle y, y \rangle$$

expand everything of $\langle x - \lambda y, x - \lambda y \rangle$ which is going to give a quadratic equation in variable λ

$$\begin{aligned} \langle x - \lambda y, x - \lambda y \rangle &= \langle x, x - \lambda y \rangle - \lambda \langle y, x - \lambda y \rangle \\ &= \langle x, x \rangle - \lambda \langle x, y \rangle - \lambda \langle y, x \rangle + \lambda^2 \langle y, y \rangle \\ &= \langle x, x \rangle - 2\lambda \langle x, y \rangle + \lambda^2 \langle y, y \rangle \end{aligned}$$

Now unless $x = \lambda y$ we have $\langle x - \lambda y, x - \lambda y \rangle > 0$ Hence the quadratic equation has no root therefore the discriminant is greater than zero.

When field is $\mathbb{C}$:

Modify the definition by

$$\langle x, y \rangle = \sum_i \overline{x_i} y_i$$

Then we still have $\langle x, x \rangle \geqslant 0$

2.3 Algorithms

Algorithm 1: what

Input: This is some input
Output: This is some output

```
/* This is a comment */
some code here;
x ← 0;
y ← 0;
if x > 5 then
  | x is greater than 5 ;                 // This is also a comment
else
  | x is less than or equal to 5;
end
foreach y in 0..5 do
  | y ← y + 1;
end
for y in 0..5 do
  | y ← y − 1;
end
while x > 5 do
  | x ← x − 1;
end
return Return something here;
```