Joint Probability Distributions Discrete case

Section 6.1

STAT/MATH 395 Spring 2020

Vincent Roulet

Lecture 5, April 8th, 2020

Ask questions via chat on Zoom
Answer quiz via PollEverywhere (username: vincentroulet)

Overview

Lecture note

- A lecture note reviewing MATH/STAT 394 is available

Homework

- 1st homework available tonight, due next Wednesday 11:59 pm
- No late homework accepted
- 1st homework is long, begin soon
- Provide **clear and detailed** answers
- One exercise chosen at random is graded to correct bad mathematical formulations
- The rest will be given points for completion

Office hours

- Answer poll to maximize availability
- For the moment, might be updated
  - Mondays 10:00 to 11:00 with T.A. Z. Yuan by Zoom
  - Fridays 11:00 to 12:00 with V. Roulet by Zoom

Overview

Previous lectures

- Probability space, probability distributions
- Probability mass function, probability density function, cumulative distribution function
- Expectation, Variance
- Various discrete and continuous random variables

This lecture

- Joint distributions discrete
- Marginal distributions
- Multinomial distribution

Answer Previous Quizzes

Exercise

If $X \sim \mathcal{N}(\mu, \sigma^2)$*, how can we choose* $a, b \in \mathbb{R}$ *s.t.* $Z = aX + b \sim \mathcal{N}(0, 1)$*?*

Answer

1. $\mathrm{Var}(Z) = a^2 \mathrm{Var}(X) = a^2\sigma^2$.
2. So to have $\mathrm{Var}(Z) = 1$, we need $a = 1/\sigma$
3. $\mathbb{E}(Z) = a\,\mathbb{E}(X) + b = a\mu + b$
4. So to have $\mathbb{E}(Z) = 0$, we need $b = -\mu/\sigma$
5. Answer was 4. i.e. $Z = \frac{X-\mu}{\sigma}$

Answer Previous Quizzes

Exercise

Waiting time of a phone call modeled by an exponential r.v.
The average waiting time is 5min.
What is the probability that the waiting time is more than 8 min?

Answer

1. $X \sim \text{Poisson}(\lambda)$
2. $\mathbb{E}(X) = \frac{1}{\lambda} = 5$ so $\lambda = 1/5$
3. $\mathbb{P}(X \geq 8) = \mathbb{P}(X > 8) = 1 - F(8) = e^{-\lambda 8} = e^{-8/5} \approx 0.20$
4. ⚠ **Typo in quiz** , none of the answers were correct

Multivariate Random Variable/Random Vector

Definition (Multivariate random variable/Random vector)

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a **multivariate random variable** or **random vector** is a vector $X = (X_1, ..., X_n)$, whose components are real-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$.

Note: Rather than speaking about the distribution of a random vector, we often speak about the joint distribution of the r.v. it is composed of

Multivariate Random Variable/Random Vector

Definition (Multivariate random variable/Random vector)

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a **multivariate random variable** or **random vector** is a vector $X = (X_1, ..., X_n)$, whose components are real-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$.

Note: Rather than speaking about the distribution of a random vector, we often speak about the joint distribution of the r.v. it is composed of

Example (Classic examples)

1. (Discrete case) Roll a die 100 times, denote $X_1, \ldots, X_6$ the number of 1, ..., 6 you got respectively, then $X = (X_1, \ldots, X_6)$ is a random vector
2. (Continuous case) Throw a dart uniformly at random on a disc, the coordinates (X, Y) of that throw form a random vector

Joint Probability Mass Function

Definition (Joint probability mass function)

Let $X_1, \ldots, X_n$ be discrete r.v. on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, their **joint probability mass function** is defined as

$$p(k_1, \ldots, k_n) = \mathbb{P}(\{X_1 = k_1\} \cap \ldots \cap \{X_n = k_n\})$$
$$\triangleq \mathbb{P}(X_1 = k_1, \ldots, X_k = k_n)$$

for any $k_1, \ldots, k_n \in X_1(\Omega) \times \ldots \times X_n(\Omega)$ (any values taken by the random vector)

Note:

- Describe all joint values of the r.v.
- We then naturally have $p(k_1, \ldots, k_n) \geq 0$ and

$$\sum_{k_1, \ldots k_n \in X_1(\Omega) \times \ldots \times X_n(\Omega)} p(k_1, \ldots, k_n) = 1$$

Joint Probability Mass Function

Example

1. Roll two dice with **4 faces**, denote
   (i) S the sum of the two dice
   (ii) Y the indicator variable that you get a pair
2. Record which outcomes lead to different values of S, Y

Joint Probability Mass Function

Example

1. Roll two dice with **4 faces**, denote
   (i) S the sum of the two dice
   (ii) Y the indicator variable that you get a pair
2. Record which outcomes lead to different values of S, Y

		Y	
		0	1
	2		(1,1)
	3	(1, 2) (2, 1)	
	4	(1, 3) (3, 1)	(2, 2)
S	5	(1, 4) (2, 3) (3, 2) (4, 1)	
	6	(2, 4) (4, 2)	(3, 3)
	7	(3, 4) (4, 3)	
	8		(4, 4)

Joint Probability Mass Function

Example

1. Roll two dice with **4 faces**, denote
   (i) S the sum of the two dice
   (ii) Y the indicator variable that you get a pair
2. Record which outcomes lead to different values of S, Y
3. Compute the corresponding joint probability mass function of S, Y

		Y	
		0	1
	2		(1,1)
	3	(1, 2) (2, 1)	
	4	(1, 3) (3, 1)	(2, 2)
S	5	(1, 4) (2, 3) (3, 2) (4, 1)	
	6	(2, 4) (4, 2)	(3, 3)
	7	(3, 4) (4, 3)	
	8		(4, 4)

Joint Probability Mass Function

Example

1. Roll two dice with **4 faces**, denote
   (i) S the sum of the two dice
   (ii) Y the indicator variable that you get a pair
2. Record which outcomes lead to different values of S, Y
3. Compute the corresponding joint probability mass function of S, Y

		Y	
		0	1
	2		(1,1)
	3	(1, 2) (2, 1)	
	4	(1, 3) (3, 1)	(2, 2)
S	5	(1, 4) (2, 3) (3, 2) (4, 1)	
	6	(2, 4) (4, 2)	(3, 3)
	7	(3, 4) (4, 3)	
	8		(4, 4)

		Y	
		0	1
	2	0	1/16
	3	1/8	0
	4	1/8	1/16
S	5	1/4	0
	6	1/8	1/16
	7	1/8	0
	8	0	1/16

Joint Probability Mass Function

Example

1. Roll two dice with **4 faces**, denote
   (i) S the sum of the two dice
   (ii) Y the indicator variable that you get a pair
2. Record which outcomes lead to different values of S, Y
3. Compute the corresponding joint probability mass function of S, Y
4. Read e.g. $\mathbb{P}(S = 4, Y = 1) = 1/16$

		Y	
		0	1
	2		(1,1)
	3	(1, 2) (2, 1)	
	4	(1, 3) (3, 1)	(2, 2)
S	5	(1, 4) (2, 3) (3, 2) (4, 1)	
	6	(2, 4) (4, 2)	(3, 3)
	7	(3, 4) (4, 3)	
	8		(4, 4)

		Y	
		0	1
	2	0	1/16
	3	1/8	0
	4	1/8	1/16
S	5	1/4	0
	6	1/8	1/16
	7	1/8	0
	8	0	1/16

Joint Probability Mass Function

Lemma

Let $g : \mathbb{R}^n \to \mathbb{R}$ and let $X_1, \dots, X_n$ be discrete r.v. on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with joint probability mass function p, then

$$\mathbb{E}[g(X_1, \dots, X_n)] = \sum_{k_1, \dots k_n \in X_1(\Omega) \times \dots \times X_n(\Omega)} g(k_1, \dots, k_n) p(k_1, \dots k_n)$$

Note: Extends naturally previous property for univariate r.v.

Example: Take $g : \mathbb{R}^n \to \mathbb{R}$ such that $g(x_1, \dots x_n) = \max_{i \in \{1, \dots, n\}} x_i$

Joint Probability Mass Function

Example

1. Roll two dices with **4 faces**, denote
   (i) S the sum of the two dices
   (ii) Y the indicator variable that you get a pair
2. Score is the sum of the dice, doubled if it is a pair

What is the average score?

		Y	
		0	1
	2	0	1/16
	3	1/8	0
	4	1/8	1/16
S	5	1/4	0
	6	1/8	1/16
	7	1/8	0
	8	0	1/16

Joint Probability Mass Function

Example

1. Roll two dices with **4 faces**, denote
   (i) S the sum of the two dices
   (ii) Y the indicator variable that you get a pair
2. Score is the sum of the dice, doubled if it is a pair

What is the average score?

		Y	
		0	1
	2	0	1/16
	3	1/8	0
	4	1/8	1/16
S	5	1/4	0
	6	1/8	1/16
	7	1/8	0
	8	0	1/16

Solution

1. The score is $g(S, Y) = S(Y + 1)$

Joint Probability Mass Function

Example

1. Roll two dices with **4 faces**, denote
   (i) S the sum of the two dices
   (ii) Y the indicator variable that you get a pair
2. Score is the sum of the dice, doubled if it is a pair

What is the average score?

		Y	
		0	1
	2	0	1/16
	3	1/8	0
	4	1/8	1/16
S	5	1/4	0
	6	1/8	1/16
	7	1/8	0
	8	0	1/16

Solution

1. The score is $g(S, Y) = S(Y + 1)$
2. The average score reads

$$\begin{aligned}
\mathbb{E}[g(S, Y)] &= \sum_{s=2}^{8} \sum_{y=0}^{1} s(y+1)p(s, y) \\
&= \sum_{s=2}^{8} sp(s, 0) + 2 \sum_{s=2}^{8} sp(s, 1) \\
&= \frac{3 + 4 + 2 \times 5 + 6 + 7}{8} + 2 \times \frac{2 + 4 + 6 + 8}{16} = 25/4 = 6.25
\end{aligned}$$

Marginal Probability Mass Function

Definition

Let $p_{X,Y}$ be the joint probability mass function of two r.v. (X, Y). The probability mass function of X is given by,

$$p_X(k) \triangleq \mathbb{P}(X = k) = \sum_{\ell \in Y(\Omega)} p_{X,Y}(k, \ell)$$

The function p_X is called the **marginal probability distribution** of X.

Proof The events $\{B_\ell = \{Y = \ell\}\}_{\ell \in Y(\Omega)}$ form a partition of Ω y definition of a discrete random variable such that

$$\mathbb{P}(X = k) = \mathbb{P}\left(\{X = k\} \cap \bigcup_{\ell=-\infty}^{+\infty} B_\ell\right) = \sum_{\ell=-\infty}^{+\infty} \mathbb{P}(X = k, Y = \ell) = \sum_{\ell \in Y(\Omega)} p_{X,Y}(k, \ell)$$

Marginal Probability Mass Function

Definition

Let p be the joint probability mass function of n discrete r.v. $X_1, \ldots X_n$. The probability mass function of X_j for $j \in \{1, \ldots, n\}$ is given by for any $k \in X_j(\Omega)$,

$$p_{X_j}(k) = \sum_{\substack{\ell_1, \ldots, \ell_{j-1}, \ell_{j+1}, \ldots, \ell_n \\ \in X_1(\Omega) \times \ldots \times X_{j-1}(\Omega) \times X_{j+1}(\Omega) \times \ldots \times X_n(\Omega)}} p(\ell_1, \ldots, \ell_{j-1}, k, \ell_{j+1}, \ldots, \ell_n)$$

The function p_{X_j} is called the **marginal probability distribution** of X_j.

Proof Denote $p_{X_{i_1}, \ldots X_{i_j}}$ the joint p.m.f. of any subset $X_{i_1}, \ldots, X_{i_j}$ of r.v. with $2 \leq j \leq n$ and $1 \leq i_1 < \ldots < i_j \leq i_n$, then naturally

$$p_{X_{i_1}, \ldots X_{i_{j-1}}}(k_{i_1}, \ldots, k_{i_{j-1}}) = \sum_{\ell_{i_j} \in X_{i_j}(\Omega)} p_{X_{i_1}, \ldots X_{i_j}}(k_{i_1}, \ldots, \ell_{i_j})$$

By applying recursively this fact we get the result.

Marginal Probability Mass Function

Previous result generalizes to the joint probability distribution of any subset. For example the joint probability of $X_1, \dots X_m$ given $m < n$ is

$$p_{X_1,\dots X_m}(k_1, \dots, k_m) = \sum_{\ell_{m+1},\dots,\ell_n \in X_{m+1}(\Omega)\times\dots\times X_n(\Omega)} p(k_1, \dots, k_m, \ell_{m+1}, \dots \ell_n)$$

Marginal Probability Mass Function

Example

1. Roll two dices with 4 faces, denote
   (i) S the sum of the two dices
   (ii) Y the indicator variable that you get a pair
2. Compute marginal distribution of Y from the joint p.m.f.

		Y	
		0	1
	2	0	1/16
	3	1/8	0
	4	1/8	1/16
S	5	1/4	0
	6	1/8	1/16
	7	1/8	0
	8	0	1/16

Marginal Probability Mass Function

Example

1. Roll two dices with 4 faces, denote
   (i) S the sum of the two dices
   (ii) Y the indicator variable that you get a pair
2. Compute marginal distribution of Y from the joint p.m.f.

		Y	
		0	1
	2	0	1/16
	3	1/8	0
	4	1/8	1/16
S	5	1/4	0
	6	1/8	1/16
	7	1/8	0
	8	0	1/16

Solution:
Sum the columns of $p(s, y)$
So you get $\mathbb{P}(Y = 1) = 4/16$ and $\mathbb{P}(Y = 0) = 12/16$

Multinomial Distribution

Motivation

Consider a trial with r possible outcomes, labeled $1, \ldots, r$. Denote p_j the probability of the out come j such that $p_1 + \ldots + p_r = 1$. Perform n independent repetitions of this trial. Denote X_j the number of times the outcome j appeared among the n trials.

What is the joint probability mass function of $(X_1, \ldots, X_r)$?

Multinomial Distribution

Motivation

Consider a trial with r possible outcomes, labeled $1, \ldots, r$. Denote p_j the probability of the out come j such that $p_1 + \ldots + p_r = 1$. Perform n independent repetitions of this trial. Denote X_j the number of times the outcome j appeared among the n trials.

What is the joint probability mass function of $(X_1, \ldots, X_r)$?

Derivation

1. Let $k_1, \ldots, k_r \in \mathbb{N}$ such that $k_1 + \ldots + k_r = n$.

Multinomial Distribution

Motivation
Consider a trial with r possible outcomes, labeled $1, \ldots, r$. Denote p_j the probability of the out come j such that $p_1 + \ldots + p_r = 1$. Perform n independent repetitions of this trial. Denote X_j the number of times the outcome j appeared among the n trials.

What is the joint probability mass function of $(X_1, \ldots, X_r)$?

Derivation

1. Let $k_1, \ldots, k_r \in \mathbb{N}$ such that $k_1 + \ldots + k_r = n$.
2. Any outcome that leads to $X_j = k_j$ for all $j \in \{1, ..., r\}$ has proba $p_1^{k_1} ... p_r^{k_r}$.

Multinomial Distribution

Motivation
Consider a trial with r possible outcomes, labeled $1, \ldots, r$. Denote p_j the probability of the out come j such that $p_1 + \ldots + p_r = 1$. Perform n independent repetitions of this trial. Denote X_j the number of times the outcome j appeared among the n trials.

What is the joint probability mass function of $(X_1, \ldots, X_r)$?

Derivation

1. Let $k_1, \ldots, k_r \in \mathbb{N}$ such that $k_1 + \ldots + k_r = n$.
2. Any outcome that leads to $X_j = k_j$ for all $j \in \{1, ..., r\}$ has proba $p_1^{k_1} ... p_r^{k_r}$.
3. The number of such outcomes is given by (in book page 392)

$$\binom{n}{k_1, \ldots, k_r} = \frac{n!}{k_1! \ldots k_r!}$$

4. Therefore we get $\mathbb{P}(X_1 = k_1, \ldots, X_r = k_r) = \binom{n}{k_1, \ldots, k_r} p_1^{k_1} ... p_r^{k_r}$

Multinomial Distribution

Definition (Multinomial distribution)

Let $n, r \in \mathbb{N}_*$, let $p_1, \ldots, p_r \in (0,1)$ s.t. $p_1 + \ldots + p_r = 1$, then a r.v. X has a **multinomial distribution** with parameters $n, r, p_1, \ldots, p_r$ if it is defined for any $k_1, \ldots k_r \in \mathbb{N}$ s.t. $k_1 + \ldots + k_r = n$ with probability

$$\mathbb{P}(X_1 = k_1, \ldots, X_r = k_r) = \binom{n}{k_1, \ldots, k_r} p_1^{k_1} \ldots p_r^{k_r}$$

We denote it $(X_1, \ldots, X_r) \sim \text{Multinom}(n, r, p_1, \ldots, p_r)$.

Note: For $r = 2$, we necessarily have $X_2 = n - X_1$ and we retrieve the binomial

Example

Roll a fair die with 6 faces 100 times
The probability that the i^{th} face appear is $p_i = 1/6$, s.t. $p_1 + \ldots + p_6 = 1$
Denote $X_1, \ldots, X_6$ the number of times face 1, ..., 6 appeared
Then $(X_1, \ldots, X_6) \sim \text{Multinom}(100, 6, \underbrace{1/6, \ldots, 1/6}_{6 \ \textit{times}})$

Exercise for next lecture

Exercise

Roll a normal die 100 times. Find the probability that among the 100 rolls, we observe exactly 22 ones, 17 fives.

Solution next lecture
Try on your own without looking at the book :)