Discrete Random variables

Statistical Computing and Empirical Methods
Unit EMATM0061, Data Science MSc

Rihuan Ke
rihuan.ke@bristol.ac.uk

Teaching Block 1, 2024

University of BRISTOL

What we will cover today

We will focus on discrete random variables and discuss the probability mass function.

We will also consider several important examples including Bernoulli and Binomial random variables.

We will study several important quantities: expectation, variance, covariance, and correlation.

We will generalise our understanding of independence to the random variable setting.

Relevant concepts

A probability space consists of a triple $(\Omega, \mathcal{E}, \mathbb{P})$, where Ω is a sample space, $\mathcal{E}$ is a well-behaved collection of events in Ω, and $\mathbb{P} : \mathcal{E} \to \mathbb{R}$ is a probability function.

Let $(\Omega, \mathcal{E}, \mathbb{P})$ be a probability space.
A pair of events $A, B \in \mathcal{E}$ are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.
A pair of events $A, B \in \mathcal{E}$ are said to be dependent if $\mathbb{P}(A \cap B) \neq \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Suppose we have a probability space $(\Omega, \mathcal{E}, \mathbb{P})$. A random variable is a mapping $X : \Omega \to \mathbb{R}$, such that for every $a, b \in \mathbb{R}$, $\{\omega \in \Omega : X(\omega) \in [a, b]\}$ is an event in $\mathcal{E}$

The distribution of a random variable X is a function given by $S \to P_X(S) := \mathbb{P}(X \in S) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in S\})$, for any $S \subseteq \mathbb{R}$ in a well-behaved collection of subsets of $\mathbb{R}$.

Discrete random variables

Support of a distribution. We say that the distribution of a random variable $X : \Omega \to \mathbb{R}$ is supported on a set $A \subseteq \mathbb{R}$ if $P_X(A) := \mathbb{P}(X \in A) = 1$.

Discrete random variables

A discrete random variable is a random variable $X : \Omega \to \mathbb{R}$ who's distribution is supported on a discrete (and hence finite or countably infinite) set $A \subseteq \mathbb{R}$

Examples.

The distribution of a Bernoulli random variable X is supported on $\{0, 1\}$, hence a discrete random variable.

The distribution of a random dice roll Z is supported on $\{1, 2, \cdots, 6\}$, hence a discrete random variable

Probability mass function

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable).

Probability mass function

The probability mass function of X is the function $p_X : \mathbb{R} \to [0,1]$ defined by

$$p_X(x) := P_X(\{x\}) = \mathbb{P}(X = x),$$

where P_X is the distribution of X.

Key features.

1. For all $x \in \mathbb{R}$, $p_X(x) \geq 0$.
2. The values of the probability mass function sum to unity $\sum_{x \in \mathbb{R}} p_X(x) = 1$.

Note: A probability mass function is a function on $\mathbb{R}$, while a probability vector (in a finite probability space) "maps" elements in Ω to their probability.

Expectation of a random variable

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable).

Expectation

The expectation $\mathbb{E}(X)$ of the random variable X is defined by $\mathbb{E}(X) := \sum_{x \in \mathbb{R}} x \cdot p_X(x)$.

We can view the expectation of a random variable as the long-run sample average obtained by repeatedly sampling independent copies of X.

The expectation is often referred to as the population average or population mean.

Variance and standard deviation

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable).

Variance and standard deviation

The variance $\mathrm{Var}(X)$ of the random variable X is defined by $\mathrm{Var}(X) := \mathbb{E}\big[(X - \mathbb{E}(X))^2\big]$.

The standard deviation of X is defined by $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.

We can view the variance of a random variable as measuring how much it typically fluctuates around its expectation $\mathbb{E}(X)$ upon repeatedly sampling independent copies of X.

The variance of a random variable is often referred to as the population variance.

The population variance and sample variance are closely connected, as we shall see.

PMF, Expectation, Variance: examples

Example. Let Z be the random variable of a dice roll.

The probability mass function is given by

$$p_Z(x) = \begin{cases} \frac{1}{6}, & \text{if } x \in \{1, 2, \cdots, 6\}. \\ 0, & \text{otherwise .} \end{cases}$$

The expectation

$$\mathbb{E}(Z) := \sum_{x \in \mathbb{R}} x \cdot p_Z(x) = \frac{1}{6}(1 + \cdots + 6) + \underbrace{\sum_{x \in \mathbb{R} \setminus \{1 \ldots 6\}} x \cdot p_Z(x)} = \frac{7}{2}.$$

The variance

$$\text{Var}(Z) := \mathbb{E}\big[(Z - \mathbb{E}(Z))^2\big] = \frac{1}{6} \sum_{x=1}^{6} \left(x - \frac{7}{2}\right)^2 = \frac{35}{12}.$$

PMF, Expectation, Variance: examples

Example. Let $X \sim \mathcal{B}(q)$ be a Bernoulli random variable.

The probability mass function is given by

$$p_X(x) = \begin{cases} 1-q & \text{if } x = 0, \\ q & \text{if } x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

The expectation

$$\mathbb{E}(X) := \sum_{x\in\mathbb{R}} x \cdot p_X(x) = (1-q) \times 0 + q \times 1 + \underbrace{\sum_{x\in\mathbb{R}\setminus\{0,1\}} x \cdot p_X(x)} = q.$$

The variance

$$\mathrm{Var}(X) := \mathbb{E}\big[(X - \mathbb{E}(X))^2\big] = \sum_{x\in R} p_X(x)\cdot(x-q)^2 = (1-q)\cdot q^2 + q\cdot(1-q)^2 = q(1-q)$$

Independent and dependent random variables

Suppose that $X_1, \cdots, X_k : \Omega \to \mathbb{R}$ are random variables, with distributions $F_{X_1}, \cdots, F_{X_k} : \mathbb{R} \to [0,1]$ defined by $F_{X_k}(x_k) := \mathbb{P}(X_k \leq x_k)$ for all x in $\mathbb{R}$.

We define the joint cumulative distribution function $F_{X_1,\cdots,X_k} : \mathbb{R}^k \to \mathbb{R}$ by

$F_{X_1,\cdots,X_k}(x_1, \cdots, x_k) = \mathbb{P}(\{X_1 \leq x_1\} \cap \cdots \cap \{X_k \leq x_k\})$ for all $(x_1, \cdots, x_k) \in \mathbb{R}^k$.

Independent random variables

We say that $X_1, \cdots X_k$ are (mutually) independent if for all $x_1, \cdots, x_k \in \mathbb{R}$, we have

$$F_{X_1,\cdots,X_k}(x_1, \cdots, x_k) = F_{X_1}(x_1) \times \cdots \times F_{X_k}(x_k)$$

Equivalently, $X_1, \cdots, X_k$ are independent if for all $x_1, \cdots, x_k \in \mathbb{R}$, the sequence of events $\{X_1 \leq x_1\}, \cdots, \{X_k \leq x_k\}$ are (mutually) independent, i.e,

$$\mathbb{P}(\{X_1 \leq x_1\} \cap \cdots \cap \{X_k \leq x_k\}) = \mathbb{P}(X_1 \leq x_1) \cdot \mathbb{P}(X_2 \leq x_2) \cdots \mathbb{P}(X_k \leq x_k)$$

We say that $X_1, \cdots, X_k$ are dependent if they are not independent.

Independent random variables: example

Example: Independence

Suppose that I roll k dice and let X_i correspond to the results of the i-th dice.

A natural assumption here is that the different dice rolls have no interaction with one another.

Hence, we can model $X_1, \cdots, X_k$ as a sequence of independent random variables.

Example: dependence

Suppose that we flip coins and let Z_j be 1 if the j-th coin was a head and 0 otherwise.

For each $i = 1, ..., k$ let $X_i = Z_1 + Z_2 + \cdots + Z_i$, the accumulated total.

The sequence $X_1, \cdots, X_k$ is a dependent sequence of random variables.

Covariance

Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are random variables.

Covariance

The covariance between X and Y is defined by

$$\mathrm{Cov}(X, Y) := \mathbb{E}\big[(X - \bar{X}) \cdot (Y - \bar{Y})\big]$$

where $\bar{X}$ and $\bar{Y}$ are the expectations of X and Y, respectively.

Recall that the variance of a random variable X is

$$\mathrm{Var}(X) = \mathbb{E}\Big[\big(X - \bar{X}\big)^2\Big].$$

Therefore $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.

The covariance between random variables is a population analogue of the sample covariance.

Correlation

We can also define the (population) correlation in terms of the (population) covariance.

Correlation

The (population) correlation is given by

$$\mathrm{Corr}(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}}$$

The correlation gives a scale-invariant quantification of the linear relation between X and Y.

Key facts:

1. If X and Y are independent random variable, then $\mathrm{Corr}(X,Y) = \mathrm{Cov}(X,Y) = 0$.

2. However, $\mathrm{Cov}(X,Y) = 0$ doesn't necessarily mean that X and Y are independent.

An alternative perspective on independence

Theorem (Independent random variables)

Let $X_1, \cdots, X_k : \Omega \to \mathbb{R}$ be a sequence of random variables. Then $X_1, \cdots, X_k$ are independent if and only if the following relationship holds for every sequence of well-behaved function (†) $f_1, f_2, \cdots, f_k$,

$$\mathbb{E}(f_1(X_1) \cdots f_k(X_k)) = \mathbb{E}(f_1(X_1)) \cdots \mathbb{E}(f_k(X_k)).$$

In particular, if X and Y are independent random variables, then $\mathrm{Cov}(X, Y) = 0$.

The variance of a linear combination of random variables

Let $X : \Omega \to \mathbb{R}$ be a random variable with distribution supported on a finite or countably infinite set (e.g. a discrete random variable).

Recall that the variance of a random variable X is $\mathrm{Var}(X) = \mathbb{E}\left[\left(X - \bar{X}\right)^2\right]$.

What is the variance of a linear combination of random variables $\sum_{i=1}^{K} \alpha_i X_i$?

Theorem (The variance of a linear combination of random variables)

Given random variables $X_1, \cdots, X_K : \Omega \to \mathbb{R}$ and $\alpha_1, \cdots, \alpha_K \in \mathbb{R}$, we have

$$\mathrm{Var}\left(\sum_{i=1}^{K} \alpha_i X_i\right) = \sum_{i=1}^{K} \alpha_i^2 \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le K} \alpha_i \alpha_j \mathrm{Cov}(X_i, X_j).$$

In particular, if $X_1, \cdots, X_k$ are independent, then

$$\mathrm{Var}\left(\sum_{i=1}^{K} \alpha_i X_i\right) = \sum_{i=1}^{K} \alpha_i^2 \mathrm{Var}(X_i)$$

Binomial distributions

We often want to model the number of successes in a sequence of (approximately) independent trials.

Examples

1. The number of red balls drawn from a bag whilst sampling with replacement.
2. The number of patients who recover following treatment in a clinical trial.
3. The number of customers who decide to buy a car following a test drive.

Binomial distributions

The Binomial distribution allows us to model the number of successes out of n independent trials, where each trial has a success probability p.

Binomial distributions

Suppose that $X_1, \cdots, X_n$ are independent random variables where each $X_i \sim \mathcal{B}(p)$ has Bernoulli distribution with $\mathbb{E}(X_i) = p$.

Then the sum $Z = X_1 + \cdots + X_n$ is a Binomial random variable with parameters n and p.

Examples

1. The number of red balls drawn from a bag whilst sampling with replacement.
2. The number of patients who recover following treatment in a clinical trial.
3. The number of customers who decide to buy a car following a test drive.

Binomial distributions

Recall that for $X_i \sim \mathcal{B}(p)$, we have $\mathbb{E}(X_i) = p$ and $\mathrm{Var}(X_i) = p(1-p)$.

Recall that given $X_i \sim \mathcal{B}(p)$, the sum $Z = X_1 + \cdots + X_n$ is a Binomial random variable with parameters n and p.

Probability mass function

$p_Z(r) := \mathbb{P}(Z = r) = \binom{n}{r} \cdot p^r \cdot (1-p)^{n-r}$ for $r \in \{0, 1, \cdots, n\}$,
$p_Z(r) = 0$ for $r \notin \{0, 1, \cdots, n\}$

Expectation

$$\mathbb{E}(Z) = \mathbb{E}(X_1 + X_2 + \cdots + X_n) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = np$$

Variance

$$\mathrm{Var}(Z) = \mathrm{Var}(X_1 + X_2 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = n \cdot p \cdot (1-p)$$

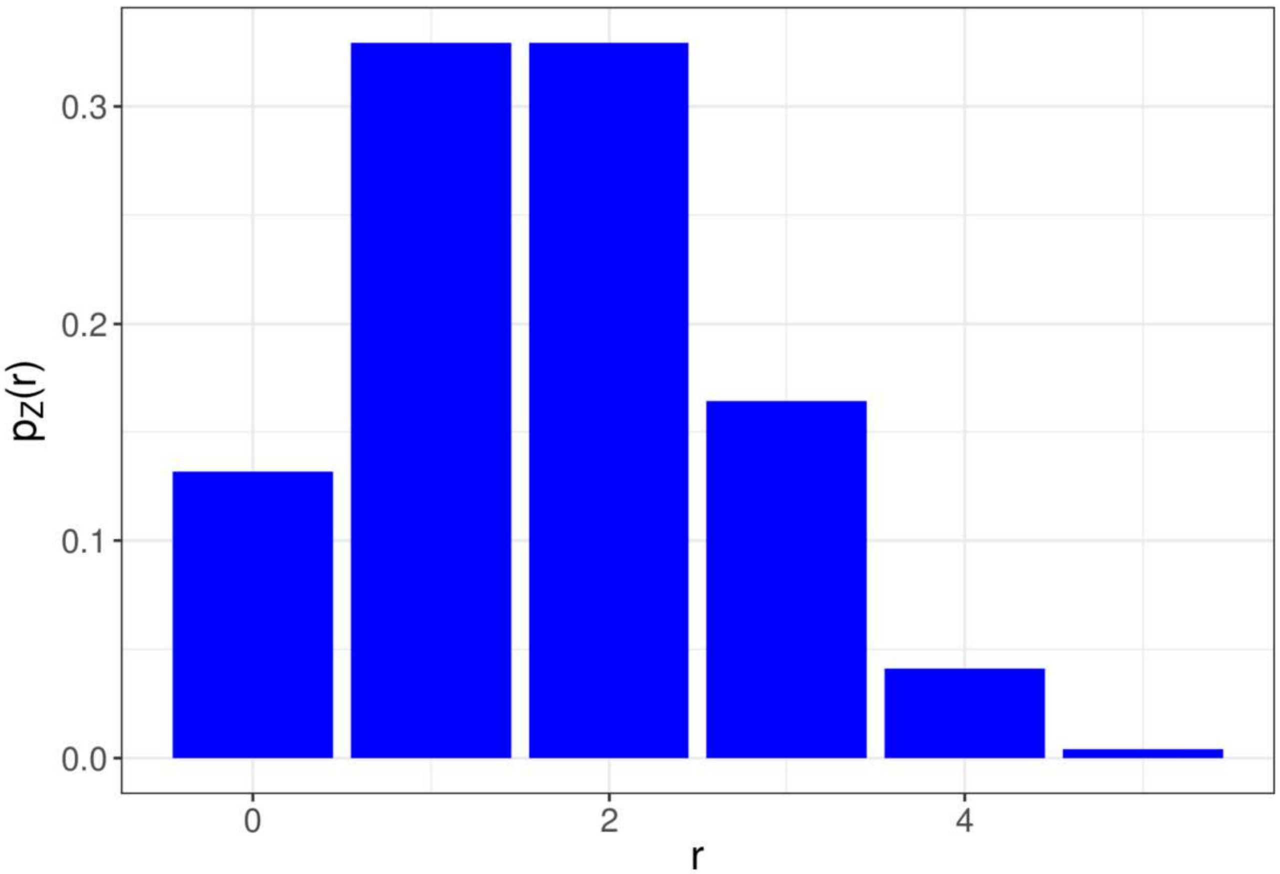
Probability mass functions of Binomial distributions

Probability mass function

$p_Z(r) := \mathbb{P}(Z = r) = \binom{n}{r} \cdot p^r \cdot (1-p)^{n-r}$ for $r \in \{0, 1, \cdots, n\}$,

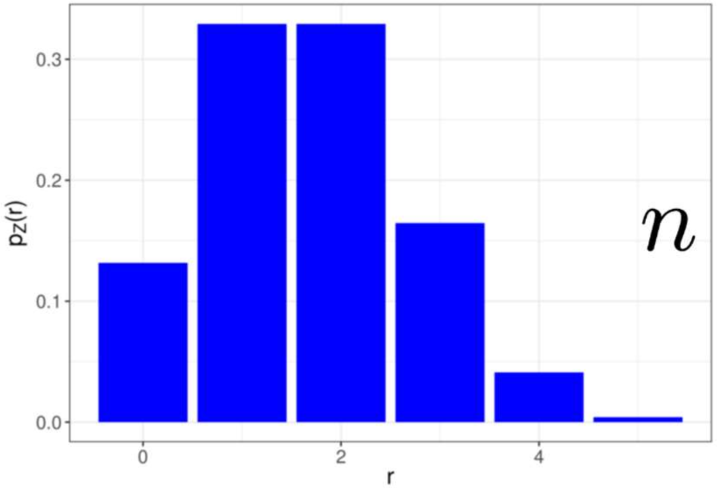
$p_Z(r) = 0$ for $r \notin \{0, 1, \cdots, n\}$

Probability mass function p_Z with $p = \frac{1}{3}$ and $n = 5$.

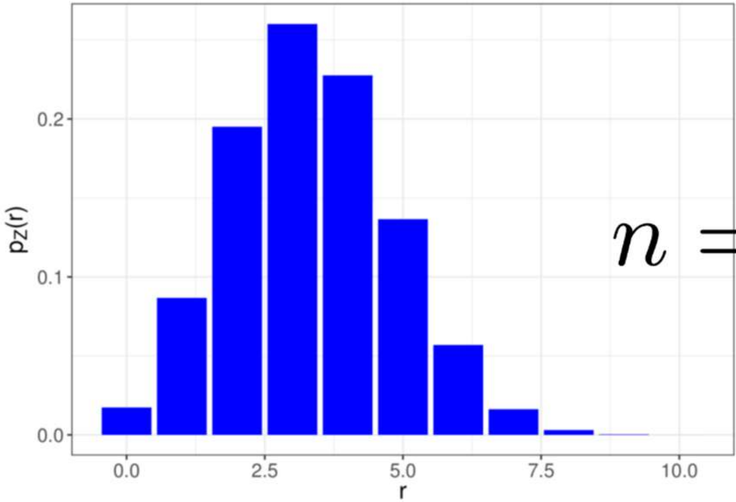


Exploring PMF for large n

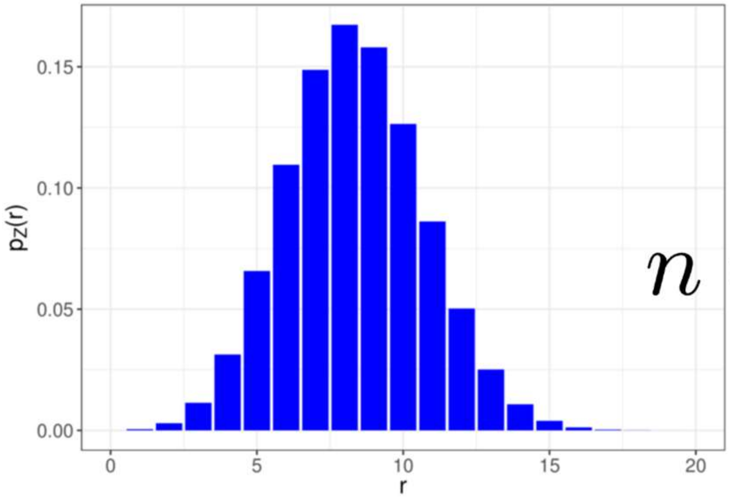
Probability mass function p_Z with $p = \frac{1}{3}$.



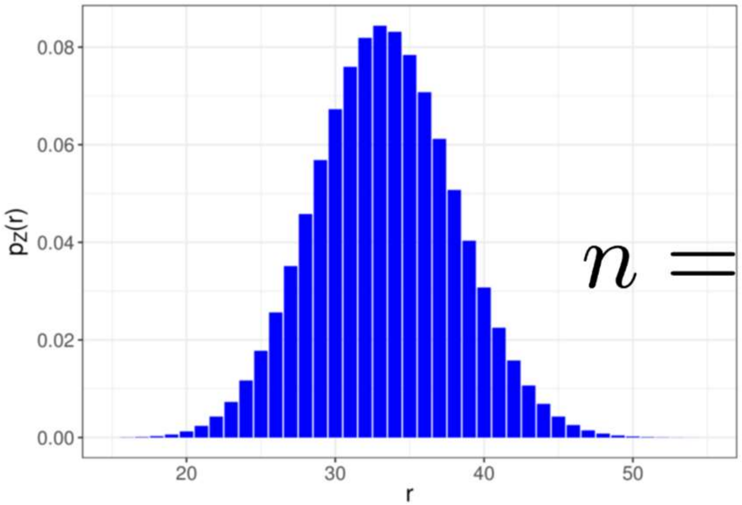
$n = 5$



$n = 10$



$n = 25$



$n = 100$

What have we covered?

We introduced the concept of a discrete random variable and discussed the probability mass function.

We discussed several important examples including Bernoulli and Binomial random variables.

We also defined the expectation, variance, covariance and correlation of random variables.

In addition, we generalized our understanding of independence from sequences of events to sequences of random variables.

University of BRISTOL

Thanks for listening!

Dr. Rihuan Ke
rihuan.ke@bristol.ac.uk

Statistical Computing and Empirical Methods
Unit EMATM0061, MSc Data Science