Discrete Random Variables

David Gerard
Many slides borrowed from Linda Collins

2017-09-28

Learning Objectives

- Random Variables
- Discrete random variables.
- Means of discrete random variables.
- Means of functions of random variables.
- Variances of discrete random variables.
- Sections 2.4, 3.3.1, 3.4 in DBC

Random Variable

- Sequence of head/tails in 4 tosses. Possible events: HHHH, HTHT, HHTT, ...
- In a clinical trial, the list of all people who had side effects: Bob, Lisa, Drake, ...
- We are usually interested in summary level data, like the number of heads in a toss or the number of people who had side effects.

random variable

A random variable is a variable whose value is a numerical outcome of a random process.

More formally, a random variable is a function from the sample space to the real numbers $\mathbb{R}$.

Example

- Suppose X is the number of heads on 4 flips of a fair coin. The domain of X is $\{HHHH, HHHT, HHTH, \ldots, TTTT\}$.
- The range of X is $\{0, 1, 2, 3, 4\}$. This is also the sample space of X.
- $X(HHHH) =$
  $X(HTTH) =$
  $X(HTHT) =$

Distribution

- We are interested in the distribution of a random variable — what values it can take and how often it takes those values.
- What is the probability of 0 heads in 4 tosses? 3 heads in 4 tosses?
- A random variable can have either a discrete or a continuous distribution.

Discrete Random Variable

A discrete random variable X has possible values that can be given in an ordered list. The probability distribution of X lists the values and their probabilities:

Value of X	x_1	x_2	x_3	$\cdots$
Probability	p_1	p_2	p_3	$\cdots$

The probabilities p_i must satisfy two requirements:

1. Every probability p_i is a number between 0 and 1.
2. $p_1 + p_2 + \cdots = 1$

Find the probability of an event by adding the probabilities p_i of the particular values x_i that make up the event.

pmf

pmf

The probability mass function (pmf) of a random variable maps the outcomes to their individual probabilities.

So $P(x_i) = p_i$, where

Value of X	x_1	x_2	x_3	$\cdots$
Probability	p_1	p_2	p_3	$\cdots$

Coin example

Let X = number of heads in 4 tosses of a fair coin.

$$\begin{aligned} P(X=1) &= P(\{HTTT, THTT, TTHT, TTTH\}) \\ &= P(HTTT) + P(THTT) + P(TTHT) + P(TTTH) \\ &= 1/16 + 1/16 + 1/16 + 1/16 \\ &= 4/16 = 1/4. \end{aligned}$$

pmf for X

We can summarize the pmf for this X in this table

x	0	1	2	3	4
$f(x)$	1/16	4/16	6/16	4/16	1/16

Note: these probabilities sum to 1:

$$\sum_x f(x) = 1/16 + 4/16 + 6/16 + 4/16 + 1/16 = 1$$

Probability Distribution for Discrete Random Variable

Can find probabilities of events using pmf:

x	0	1	2	3	4
$f(x)$	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$

$P(A) = P(X \in A)$

$A =$ Event of 3 heads

$P(A) = P(X = 3) = 4/16$

$B =$ Event that all 4 flips result in either all heads or all tails

$$P(B) = P(X = 0 \cup X = 4) = P(X = 0) + P(X = 4) = \frac{1}{16} + \frac{1}{16} = \frac{2}{16}$$

$C =$ Event that at most one of the 4 flips is heads

$P(C) = P(X \geq 3) = P(X = 3 \cup X = 4) = 4/16 + 1/16 = 5/16$

The Mean

The same way we try to describe the center and spread of data, we often want to describe the center and spread of the distribution of a random variable.

Mean (Expected Value)

Suppose X is a discrete random variable, then

$$\begin{aligned} \text{mean of } X &= E[X] \\ &= \sum_{\text{all } x} xP(X = x) \\ &= \sum_{\text{all } x} xf(x) \\ &= \mu \end{aligned}$$

Mean Intuition I: Benford's Law

- The first digits of numbers in legitimate financial records often follows Benford's law.

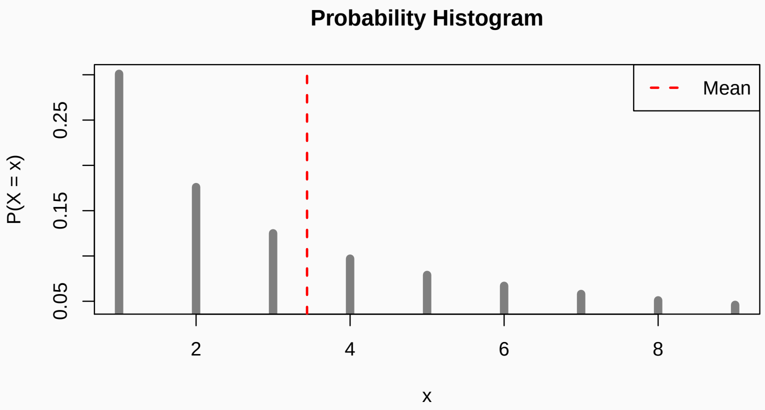
1st Digit	1	2	3	4	5	6	7	8	9
Prob	0.30	0.18	0.13	0.1	0.08	0.07	0.06	0.05	0.05

- Extreme deviations from these probabilities can alert investigators to fraud.

Mean Intuition II: Calculate Mean

$$\begin{aligned} E[X] &= 1 \times 0.30 + 2 \times 0.18 + 3 \times 0.13 + 4 \times 0.1 \\ &+ 5 \times 0.08 + 6 \times 0.07 + 7 \times 0.06 + 8 \times 0.05 + 9 \times 0.05 \\ &= 3.441 \end{aligned}$$

Mean Intuition III: Probability Histogram and Center of Mass



Mean Intuition IV: Average when simulate a lot of data

Suppose we chose 100 digits at random using the distribution from Benford's Law

Then we would expect about 30 1's, 18 2's, 13 3's, etc...

If we take an average of the values we expect, we get:

$$
\begin{aligned}
&\frac{30 \times 1 + 18 \times 2 + 13 \times 3 + \cdots + 5 \times 9}{100} \\
&= \frac{30}{100} \times 1 + \frac{18}{100} \times 2 + \frac{13}{100} \times 3 + \cdots + \frac{5}{100} \times 9 \\
&= 0.3 \times 1 + 0.18 \times 1 + 0.13 \times 1 + \cdots + 0.05 \times 9 \\
&= 3.441.
\end{aligned}
$$

Notice that we empirically derived the formula for expected value that we used earlier.

Mean Intuition V: Average when simulate a lot of data

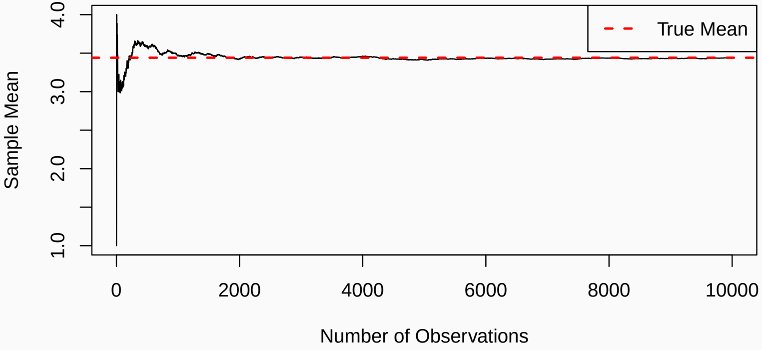
Recall: most samples of 100 digits won't be exactly equal to 3.441. But if we took a sample of MANY MANY digits, we could get pretty close:

```
ben <- c(0.301, 0.176, 0.125, 0.097,
         0.079, 0.067, 0.058, 0.051, 0.046)
sample_digits <- sample(x = 1:9, size = 10000,
                        prob = ben, replace = TRUE)
mean(sample_digits)

[1] 3.443
```

Recall that the mean is 3.441

Mean Intuition VI: Average when simulate a lot of data



So $E[X] =$ average when we take a HUGE sample.

Functions of a random variable i

If $X : S \to \mathbb{R}$ is a random variable, and $g : \mathbb{R} \to \mathbb{R}$ is a function then $Y = g(X)$ is also a random variable with range $R_Y = g(R_X)$.

Example: For Benford's law example take $g(x) = \begin{cases} 1 & \text{if } x \leq 5 \\ 0 & \text{if } x > 5 \end{cases}$.

Then $Y = g(X)$ has range $R_Y = \{1, 0\}$ and distribution
$f_Y(1) = P(Y = 1) = P(X \in \{1, 2, 3, 4, 5\}) = 0.778$
$f_Y(0) = P(Y = 0) = P(X \in \{6, 7, 8, 9\}) = 0.222$.

So we can compute the mean: $E(Y) = 1 \cdot 0.778 + 0 \cdot 0.222 = 0.778$

Sometimes it is difficult to compute the new distribution of Y.

It's still easy to compute $E(Y)$ in terms of the distribution of X.

Functions of random variables ii

Law of unconscious statistican

Suppose we know the pmf of X and $Y = g(X)$, then

$$E(Y) = Eg(X) = \sum_{k=1}^{K} g(x_k) f_X(x_k).$$

Functions of random variables iii

1st Digit	1	2	3	4	5	6	7	8	9
Prob	0.30	0.18	0.13	0.1	0.08	0.07	0.06	0.05	0.05

So in our Benford's law example we compute

$$\begin{aligned} E(Y) =& g(1) \cdot 0.301 + g(2) \cdot 0.176 + g(3) \cdot 0.125 \\ & + g(4) \cdot 0.097 + g(5) \cdot 0.079 + g(6) \cdot 0.067 \\ & + g(7) \cdot 0.058 + g(8) \cdot 0.051 + g(9) \cdot 0.046 \\ =& 1 \cdot 0.301 + 1 \cdot 0.176 + 1 \cdot 0.125 \\ & + 1 \cdot 0.097 + 1 \cdot 0.079 + 0 \cdot 0.067 \\ & + 0 \cdot 0.058 + 0 \cdot 0.051 + 0 \cdot 0.046 \\ =& 0.778. \end{aligned}$$

Parameters: Describing the Spread of a r.v.

One measure of spread a probability distribution is **variance**.

Variance: The variance of a probability distribution
is the average squared distance from the mean.

Variance

Variance of $X = Var(X) = \sigma^2 =$ "sigma squared" $= E[(X - \mu)^2]$

Variance is in squared units.
Take the square root to determine the **standard deviation.**

Standard Deviation of $X = \sqrt{Var(X)} = \sigma =$ "sigma" $= SD(X)$

One standard deviation is roughly
the "typical" distance of outcomes from the mean.

A Formula Useful for Calculating Variance of a r.v.

This formula is not intuitive! Instead, think of variance as "the average squared distance of outcomes from their mean."

However, to make calculation of variance easier, we can show that

$$Var(X) = E[(X-\mu)^2] = E(X^2) - (\mu)^2$$

$$E[(X-\mu)^2] = \sum_x (x-\mu)^2 \, f(x)$$

$$=$$

$$=$$

$$\underbrace{\qquad\qquad} \qquad \underbrace{\qquad\qquad} \qquad \underbrace{\qquad\qquad}$$

$$= \qquad\qquad =$$

A Formula Useful for Calculating Variance of a r.v.

This formula is not intuitive! Instead, think of variance as "the average squared distance of outcomes from their mean."

However, to make calculation of variance easier, we can show that

$$Var(X) = E[(X-\mu)^2] = E(X^2) - (\mu)^2$$

$$\begin{aligned} E[(X-\mu)^2] &= \sum_x (x-\mu)^2 \, f(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2) \, f(x) \\ &= \\ & \quad \underbrace{\qquad\qquad\qquad} \qquad \underbrace{\qquad\qquad\qquad} \qquad \underbrace{\qquad\qquad} \\ &= \qquad\qquad\qquad\qquad = \end{aligned}$$

A Formula Useful for Calculating Variance of a r.v.

This formula is not intuitive! Instead, think of variance as "the average squared distance of outcomes from their mean."

However, to make calculation of variance easier, we can show that

$$Var(X) = E[(X-\mu)^2] = E(X^2) - (\mu)^2$$

$$\begin{aligned}
E[(X-\mu)^2] &= \sum_x (x-\mu)^2 \, f(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2) \, f(x) \\
&= \underbrace{\sum_x x^2 \, f(x)}_{} - 2\mu \underbrace{\sum_x x \, f(x)}_{} + \mu^2 \underbrace{\sum_x f(x)}_{} \\
&= \qquad\qquad\qquad\qquad =
\end{aligned}$$

A Formula Useful for Calculating Variance of a r.v.

This formula is not intuitive! Instead, think of variance as "the average squared distance of outcomes from their mean."

However, to make calculation of variance easier, we can show that

$$Var(X) = E[(X-\mu)^2] = E(X^2) - (\mu)^2$$

$$\begin{aligned} E[(X-\mu)^2] &= \sum_x (x-\mu)^2\, f(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2)\, f(x) \\ &= \underbrace{\sum_x x^2\, f(x)}_{=E(X^2)} - 2\mu \underbrace{\sum_x x\, f(x)}_{} + \mu^2 \underbrace{\sum_x f(x)}_{} \\ &= \qquad\qquad\qquad\qquad = \end{aligned}$$

A Formula Useful for Calculating Variance of a r.v.

This formula is not intuitive! Instead, think of variance as "the average squared distance of outcomes from their mean."

However, to make calculation of variance easier, we can show that

$$Var(X) = E[(X - \mu)^2] = E(X^2) - (\mu)^2$$

$$\begin{aligned}
E[(X-\mu)^2] &= \sum_x (x-\mu)^2 \, f(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2) \, f(x) \\
&= \underbrace{\sum_x x^2 \, f(x)}_{= E(X^2)} - 2\mu \underbrace{\sum_x x \, f(x)}_{= \mu} + \mu^2 \underbrace{\sum_x f(x)}_{} \\
&= \qquad\qquad\qquad =
\end{aligned}$$

A Formula Useful for Calculating Variance of a r.v.

This formula is not intuitive! Instead, think of variance as "the average squared distance of outcomes from their mean."

However, to make calculation of variance easier, we can show that

$$Var(X) = E[(X-\mu)^2] = E(X^2) - (\mu)^2$$

$$\begin{aligned}
E[(X-\mu)^2] &= \sum_x (x-\mu)^2 \, f(x) \\
&= \sum_x (x^2 - 2\mu x + \mu^2) \, f(x) \\
&= \underbrace{\sum_x x^2 \, f(x)}_{=E(X^2)} - 2\mu \underbrace{\sum_x x \, f(x)}_{=\mu} + \mu^2 \underbrace{\sum_x f(x)}_{=1} \\
&= \qquad\qquad\qquad =
\end{aligned}$$

A Formula Useful for Calculating Variance of a r.v.

This formula is not intuitive! Instead, think of variance as "the average squared distance of outcomes from their mean."

However, to make calculation of variance easier, we can show that

$$Var(X) = E[(X-\mu)^2] = E(X^2) - (\mu)^2$$

$$\begin{aligned} E[(X-\mu)^2] &= \sum_x (x-\mu)^2 \, f(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2) \, f(x) \\ &= \underbrace{\sum_x x^2 \, f(x)}_{=E(X^2)} - 2\mu \underbrace{\sum_x x \, f(x)}_{=\mu} + \mu^2 \underbrace{\sum_x f(x)}_{=1} \\ &= E(X^2) - 2\mu^2 + \mu^2 \qquad = \end{aligned}$$

A Formula Useful for Calculating Variance of a r.v.

This formula is not intuitive! Instead, think of variance as "the average squared distance of outcomes from their mean."

However, to make calculation of variance easier, we can show that

$$Var(X) = E[(X-\mu)^2] = E(X^2) - (\mu)^2$$

$$\begin{aligned} E[(X-\mu)^2] &= \sum_x (x-\mu)^2\, f(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2)\, f(x) \\ &= \underbrace{\sum_x x^2\, f(x)}_{=E(X^2)} - 2\mu \underbrace{\sum_x x\, f(x)}_{=\mu} + \mu^2 \underbrace{\sum_x f(x)}_{=1} \\ &= E(X^2) - 2\mu^2 + \mu^2 \qquad = E(X^2) - \mu^2 \end{aligned}$$