Random Variables

Sumanta Basu

Announcement

Make-up Prelim 1: Friday October 4 (time and place TBA)

Email Prof. Basu by this weekend if you need to take make-up prelim 1

iClickers and attendance in lectures/labs are not mandatory [**unless you are auditing this class**], but strongly encouraged. Class participation (through iClickers and other activities) will be used to calculate bonus points (upto 4%) in your course totals.

Reading: Textbook Section 2.4

Recommended Exercise: 2.33, 2.37, 2.39, 2.41

Random variables

Definition

Random variable:

> A numerical summary of a random outcome

Example

- Experiment: Toss **3 fair coins**
- Sample space?

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

- Each equally likely,

$$P(\text{each outcome}) = \frac{1}{8}$$

(*assuming what?*)

- **Random variable:** $X =$ number of heads.

$$P(X = 2) = ?$$

A closer look at X

$$X = \text{number of heads}$$

- before actually doing experiment, do not know which value of X will be seen
- the **possible values** of X are known in advance

$$X = \begin{cases} 3 & \text{if } HHH \\ 2 & \text{if } HHT \text{ or } HTH \text{ or } THH \\ 1 & \text{if } HTT \text{ or } THT \text{ or } TTH \\ 0 & \text{if } TTT \end{cases}$$

> X is a *function* that assigns a number to each possible outcome in Ω.

$$X = \begin{cases} 3 & \text{if } HHH \\ 2 & \text{if } HHT \text{ or } HTH \text{ or } THH \\ 1 & \text{if } HTT \text{ or } THT \text{ or } TTH \\ 0 & \text{if } TTT \end{cases}$$

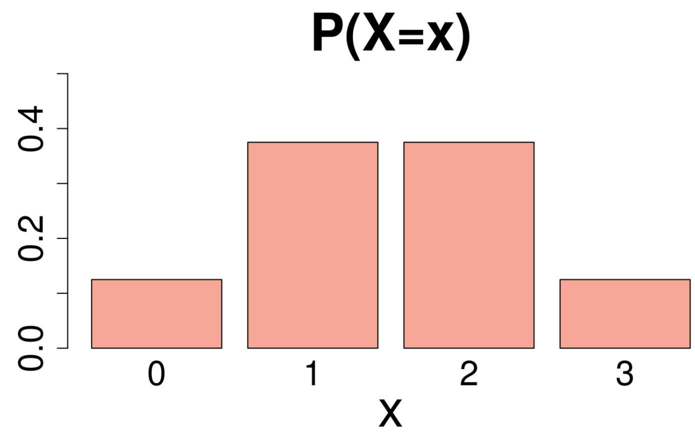
Event of interest: $X =$ *some value*

$$P(X = x) = \begin{cases} 1/8 & \text{if } x = 3 \\ 3/8 & \text{if } x = 2 \\ 3/8 & \text{if } x = 1 \\ 1/8 & \text{if } x = 0 \end{cases}$$

This is called the **probability mass function** of X.

Probability mass function

$$P(X = x) = \begin{cases} 1/8 & \text{if } x = 3 \\ 3/8 & \text{if } x = 2 \\ 3/8 & \text{if } x = 1 \\ 1/8 & \text{if } x = 0 \end{cases}$$



Similar to a density histogram…
but these are actual probabilities
rather than observed frequencies.

What about a general coin?

This still holds:

$$X = \begin{cases} 3 & \text{if } HHH \\ 2 & \text{if } HHT \text{ or } HTH \text{ or } THH \\ 1 & \text{if } HTT \text{ or } THT \text{ or } TTH \\ 0 & \text{if } TTT \end{cases}$$

But if $P(H) = p$

$$P(X = x) = \begin{cases} ? & \text{if } x = 3 \\ ? & \text{if } x = 2 \\ ? & \text{if } x = 1 \\ ? & \text{if } x = 0 \end{cases}$$

Assumptions we make

- result of each toss is **independent**
- $P(H) = p$ for each toss

Calculation

$$P(X = 2) = P(HHT \text{ or } HTH \text{ or } THH)$$

1. They are **disjoint**, so

$$P(X = 2) = P(HHT) + P(HTH) + P(THH)$$

1. Coin tosses are **independent**, so

$$P(HHT) = P(\text{first} = H)P(\text{second} = H)P(\text{third} = T)$$

1. Each flip has identical probability of heads, p:

$$P(HHT) = p^2(1 - p)$$

2. Putting this together:

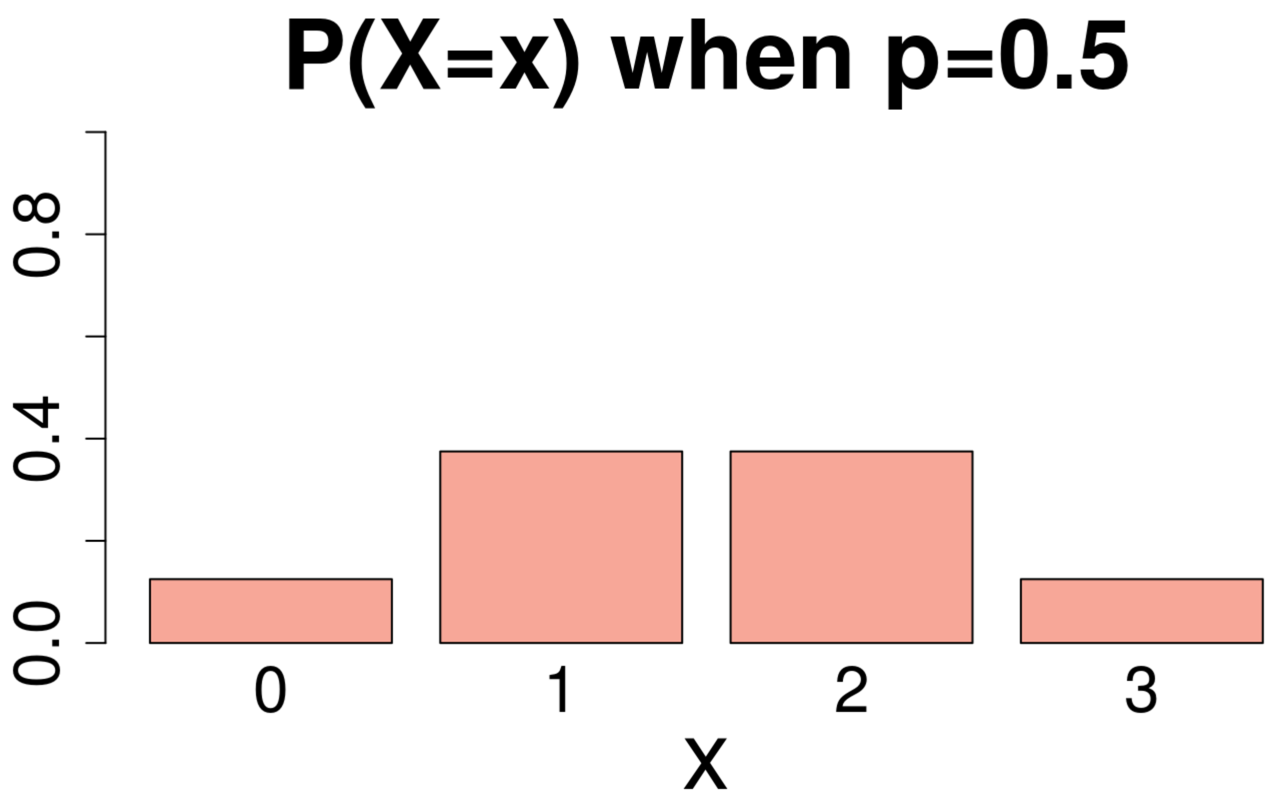
$$P(X = 2) = p^2(1 - p) + p(1 - p)p + (1 - p)p^2 = 3p^2(1 - p)$$

Probability mass function

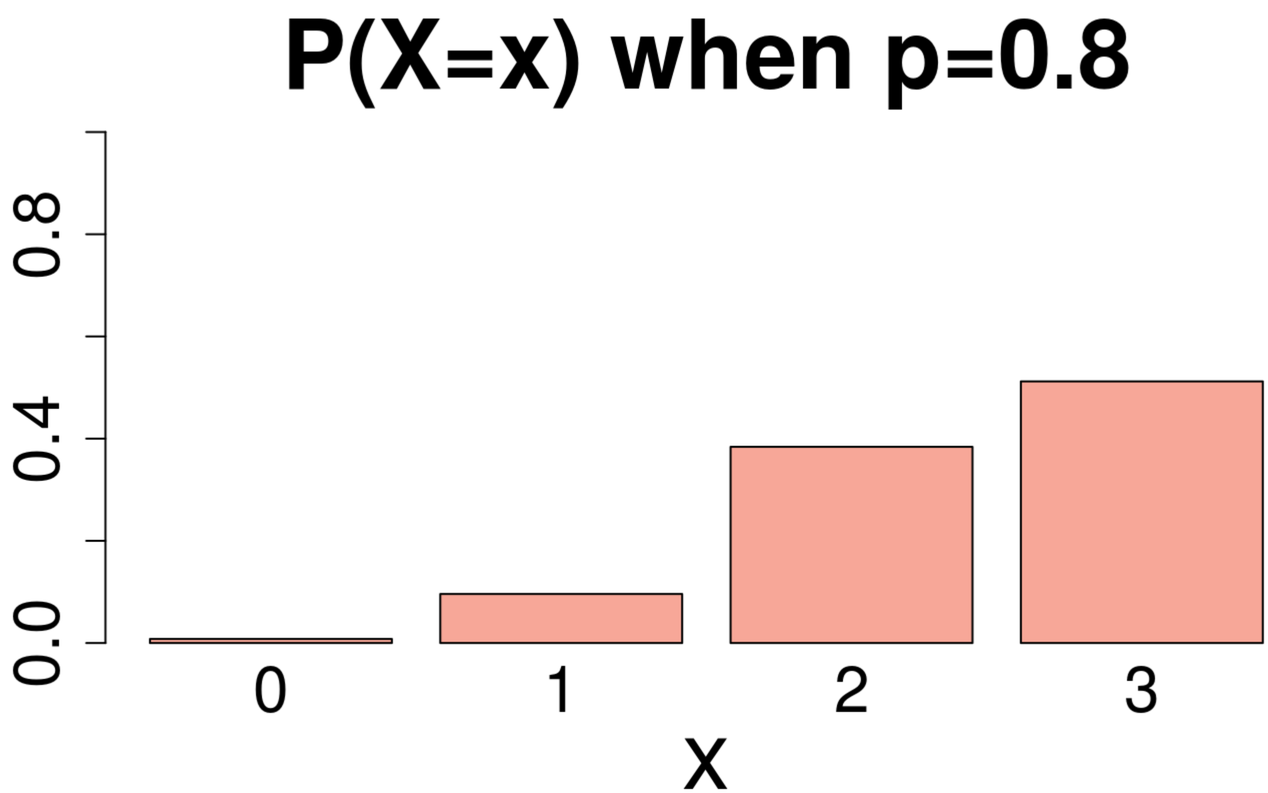
$$P(X = x) = \begin{cases} p^3 & \text{if } x = 3 \\ 3p^2(1-p) & \text{if } x = 2 \\ 3p(1-p)^2 & \text{if } x = 1 \\ (1-p)^3 & \text{if } x = 0 \end{cases}$$

Does this agree with $p = 1/2$ case?

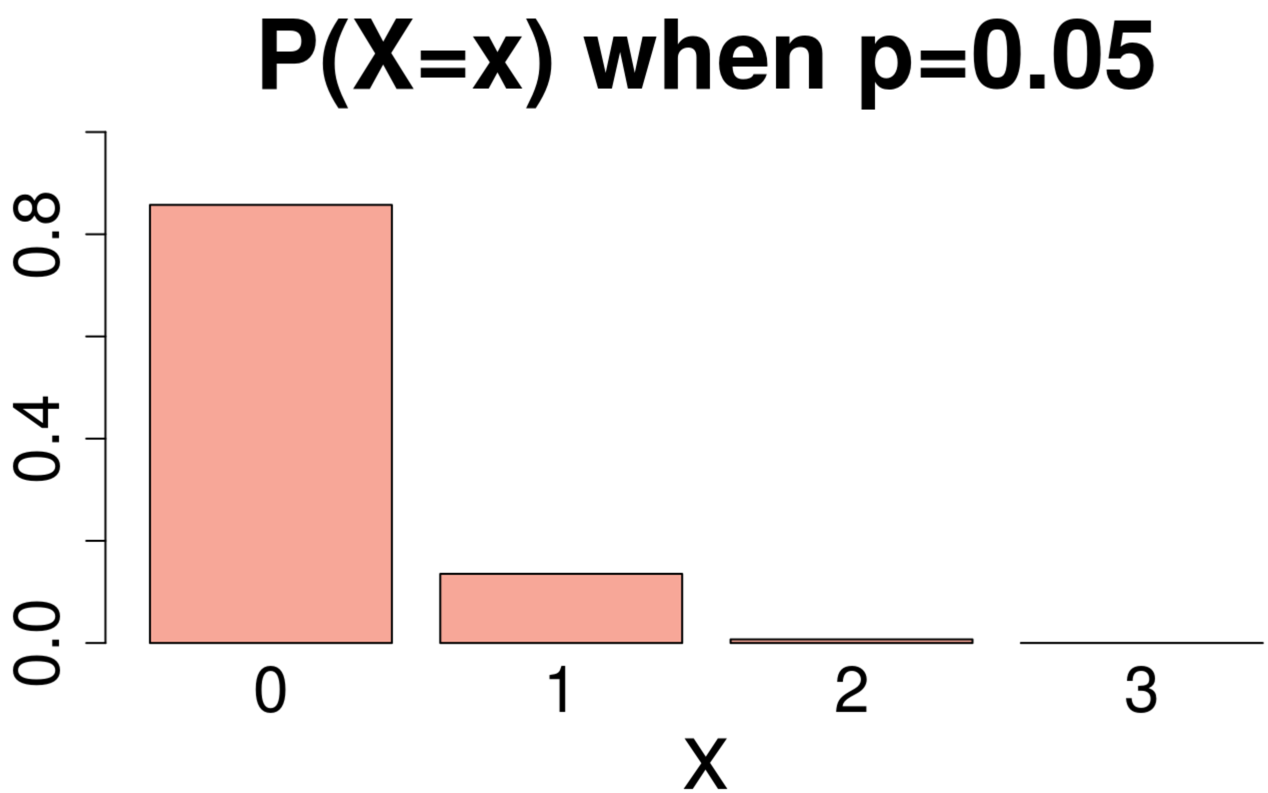
Probability mass function



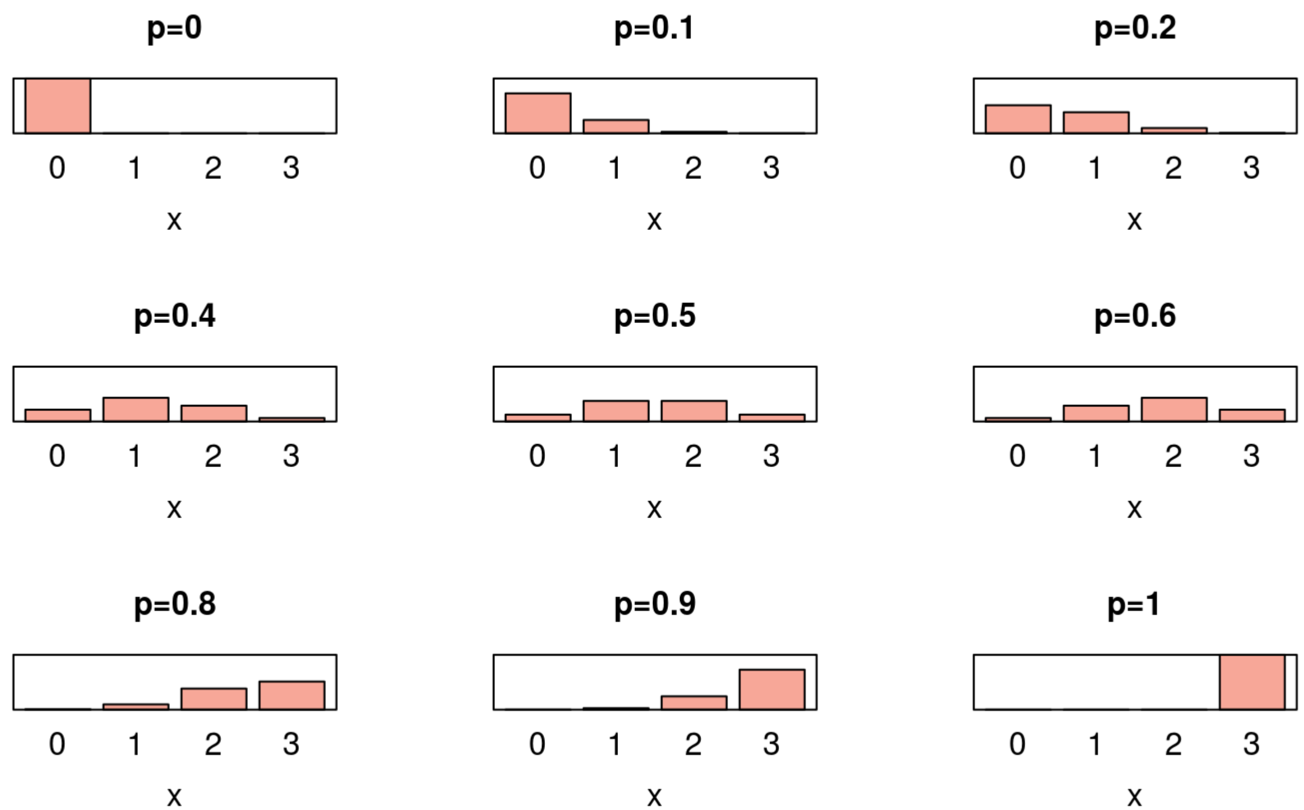
Probability mass function



Probability mass function



Probability mass function



Discrete Random Variable

Random variable: *a numerical summary of a random outcome*

Discrete random variable: *random variable that takes on a "countable" number of possible values*

$$x_1, x_2, x_3, x_4, \ldots$$

e.g., X = number of heads in n tosses of a coin

$$0, 1, 2, \ldots, n$$

Probability mass function

A discrete random variable is described by

1. its possible set of **values** $x_1, x_2, \ldots$

2. the **probability** of each such value: $p_1, p_2, \ldots$ where

$$p_j = P(X = x_j)$$

called the **probability mass function** (or **pmf**) of X

Note:

$$p_j \geq 0 \text{ and } \sum_j p_j = 1$$

Example: Toss coin 3 times

Fair coin

j	1	2	3	4
x_j	0	1	2	3
p_j	1/8	3/8	3/8	1/8

General coin

j	1	2	3	4
x_j	0	1	2	3
p_j	p^3	$3p^2(1-p)$	$3p(1-p)^2$	$(1-p)^3$

Toss two dice

Define X = sum of dice

What is pmf?

- draw picture

Realization of a random variable

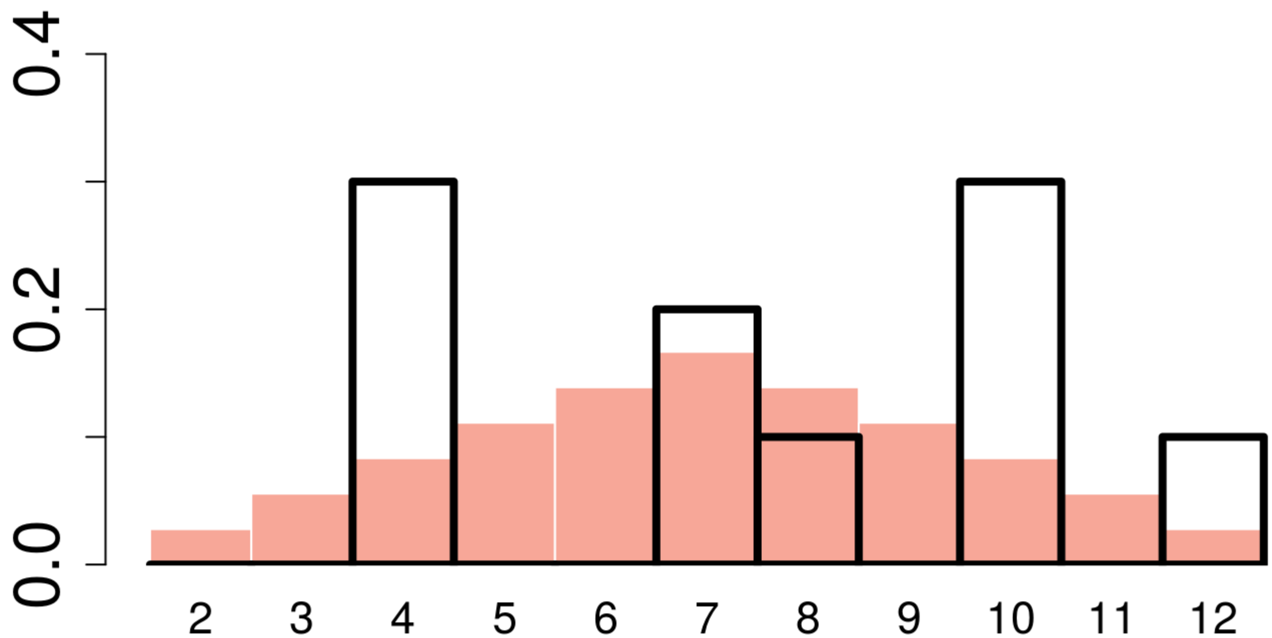
- random variable X is an abstract concept, described by its **probability distribution**
- represents expectations before experiment
- when we roll two dice, we get a specific outcome, e.g., (6,6) and a specific value, 12.
- 12 is called a **realization** of the random variable.
- repeating experiment fifty times gives fifty realizations:

```
## 3 8 8 7 10 3 5 11 10 6 10 4 8 6 6 9 7 6 7 9 11 9 9 7 7 7 7 8 5 2 8 11 8 10 2 6 1
```

- Recall: $P(X = 7)$ tells us about **long-run frequency** of 7.

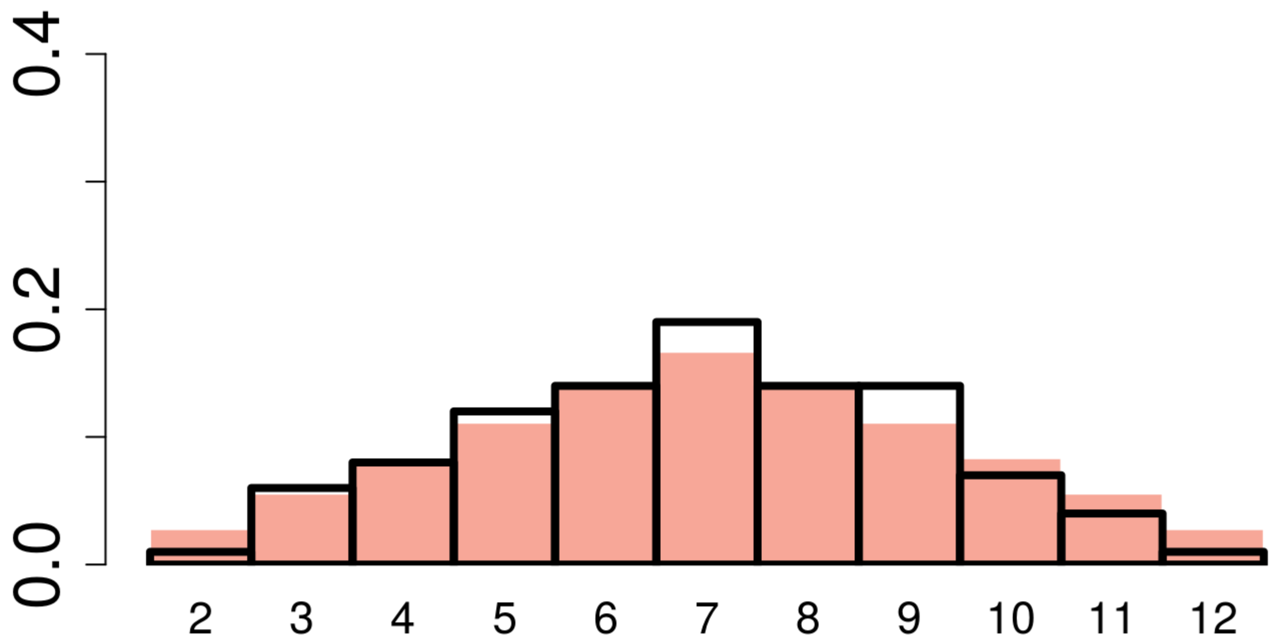
PMF as long-run frequency

Suppose I ask 10 of you to roll 2 dice…



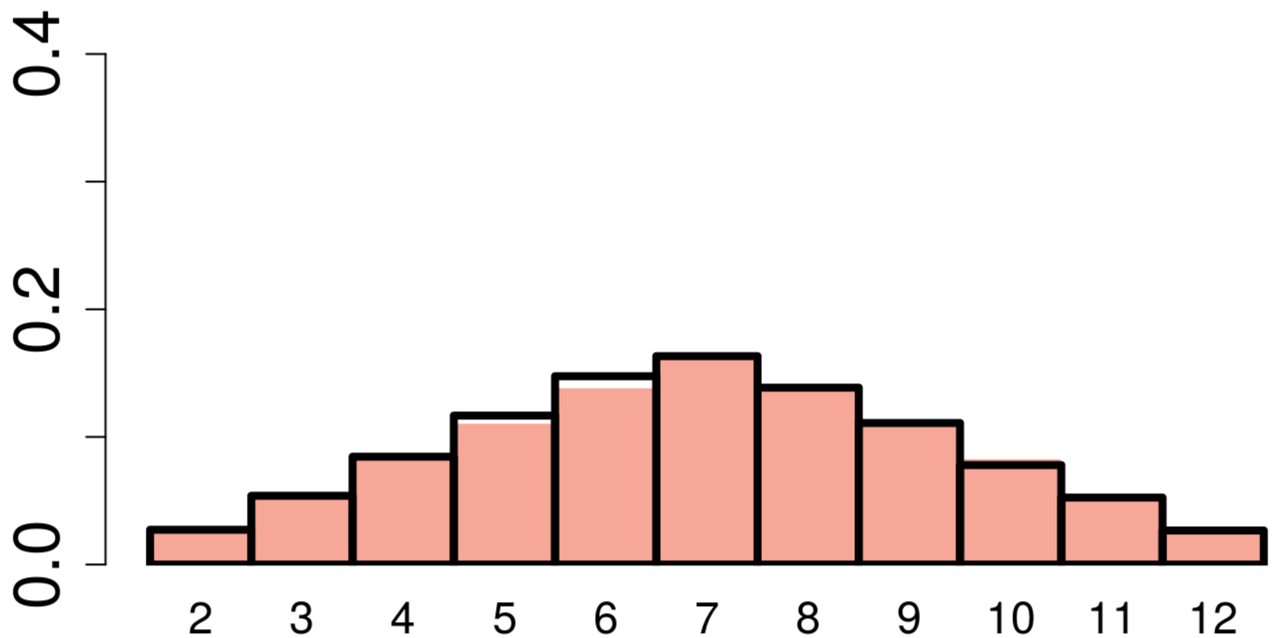
PMF as long-run frequency

Suppose I ask 100 of you to roll 2 dice…



PMF as long-run frequency

Suppose I ask 10000 of you to roll 2 dice…



Terminology

- $P(X = x)$ - probability mass function or pmf
- $P(X \leq x)$ - cumulative distribution function or cdf

Example:

- $P(X = 7)$?
- $P(X \leq 6)$?
- $P(X > 6)$?
- $P(3 \leq X \leq 6)$?

Summarizing a probability distribution

The **pmf** is like an **idealized probability histogram**

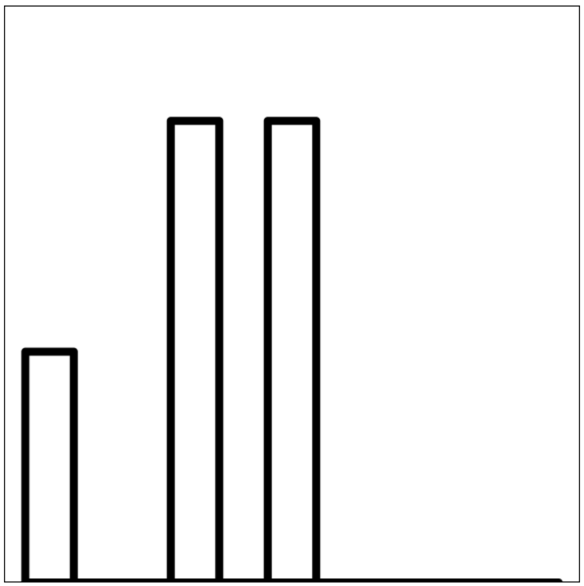
- except the *histogram* was based on a *sample* (e.g., 10 rolls)
- the *pmf* represents the "**population**" (of throwing 2 dice)

It also can be summarized

Sample vs. Population

$7, 5, 2, 5, 7$

histogram

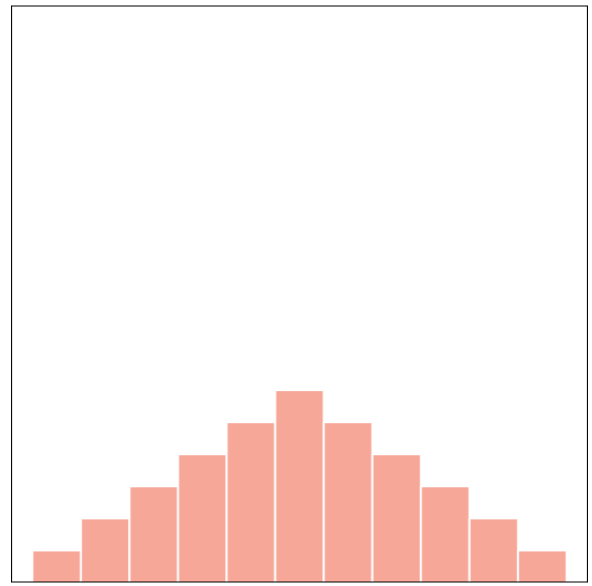


Sample mean, $\bar{x}$

Sample variance, s^2

X

pmf



Expected value, $E(X)$

Variance, $\mathrm{Var}(X)$

Expected value

For a discrete random variable X, the **expected value** (or **mean**) is

$$\mu = E(X) = \sum_i x_i P(X = x_i)$$

- weighted average of the possible values, weighted by how likely
- measures "center" of pmf

Variance

For a discrete random variable X, the **variance** is

$$\sigma^2 = \mathrm{Var}(X) = \sum_i (x_i - \mu)^2 P(X = x_i)$$

- weighted average of the squared distances from each possible value to the mean
- measures spread of pmf
- σ is called the **standard deviation**
- return to picture

Properties of Random Variables (RV)

Expectation or Mean:

For any RV X,

- $E(cX) = cE(X)$, for any constant c,
- $E(X + c) = E(X) + c$, for any constant c,

For any two RVs X and Y, and constants a, b,

- $E(aX + bY) = aE(X) + bE(Y)$
- e.g., $E(2X - 3Y) = 2E(X) - 3E(Y)$

In particular, $E(X + Y) = E(X) + E(Y)$

Properties of Random Variables (RV)

Variance:

$Var(c + X) = Var(X)$, for any constant c

- Heuristic: constants do not contribute to variation

$Var(cX) = c^2 Var(X)$, for any constant c

- e.g., $Var(-2X) = 4Var(X)$
- Why c^2 and not c??

$SD(cX) = |c| SD(X)$, for any constant c

- e.g., $SD(-2X) = 2SD(X)$

For two independent RVs X, Y, and constants a, b,

- $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y)$

Properties of Random Variables (RV)

More generally,

$$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$$

And for independent RVs $X_1, X_2, \ldots, X_n$,

$$Var(X_1 + X_2 + \ldots + X_n) = Var(X_1) + Var(X_2) + \ldots + Var(X_n)$$

In other words,

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$$

and for independent RVs $X_1, X_2, \ldots, X_n$,

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$$

Prestige Data

```
library(car)
head(Prestige, n=10)
```

```
##                      education income women prestige census type
## gov.administrators       13.11  12351 11.16     68.8   1113 prof
## general.managers         12.26  25879  4.02     69.1   1130 prof
## accountants              12.77   9271 15.70     63.4   1171 prof
## purchasing.officers      11.42   8865  9.11     56.8   1175 prof
## chemists                 14.62   8403 11.68     73.5   2111 prof
## physicists               15.64  11030  5.13     77.6   2113 prof
## biologists               15.09   8258 25.65     72.6   2133 prof
## architects               15.44  14163  2.69     78.1   2141 prof
## civil.engineers          14.52  11377  1.03     73.1   2143 prof
## mining.engineers         14.64  11023  0.94     68.8   2153 prof
```

Prestige Data

```
# average education (in years)
mean(Prestige$education)
```

```
## [1] 10.73804
```

```
# how much does it vary?
var(Prestige$education)
```

```
## [1] 7.444408
```

```
sd(Prestige$education)
```

```
## [1] 2.728444
```

What if we measured education in months?

```
Prestige$education_month <- 12*Prestige$education
var(Prestige$education_month)
```

```
## [1] 1071.995
```

```
12^2*var(Prestige$education)
```

```
## [1] 1071.995
```

```
sd(Prestige$education_month)
```

```
## [1] 32.74133
```

Prestige Data

```r
# create a new score: education + women
Prestige$edu.women =
  Prestige$education + Prestige$women

# how much does it vary?
var(Prestige$edu.women)
```

```
## [1] 1024.624
```

```r
var(Prestige$education)+
  var(Prestige$women)
```

```
## [1] 1013.916
```

Two expressions are very close!

```r
# create another new score: education + prestige
Prestige$edu.prestige =
  Prestige$education + Prestige$prestige

# how much does it vary?
var(Prestige$edu.prestige)
```

```
## [1] 383.2559
```

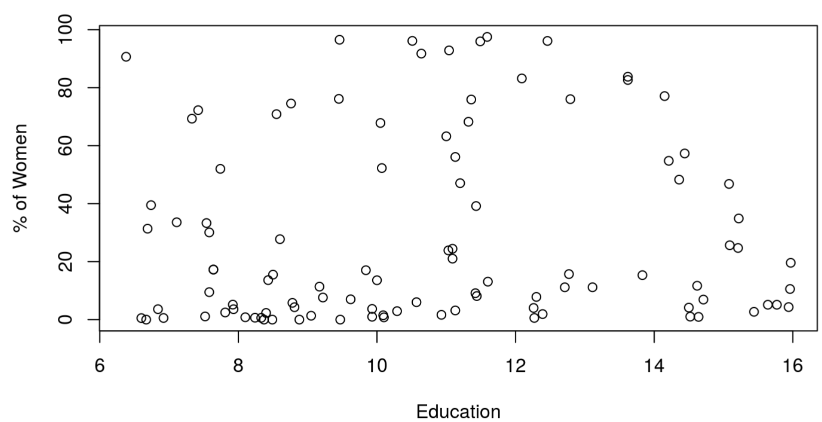
```r
var(Prestige$education)+
var(Prestige$prestige)
```

```
## [1] 303.4387
```

They are quite different!

Prestige Data

```r
plot(Prestige$education,
     Prestige$women,
     xlab = 'Education',
     ylab = '% of Women')
```

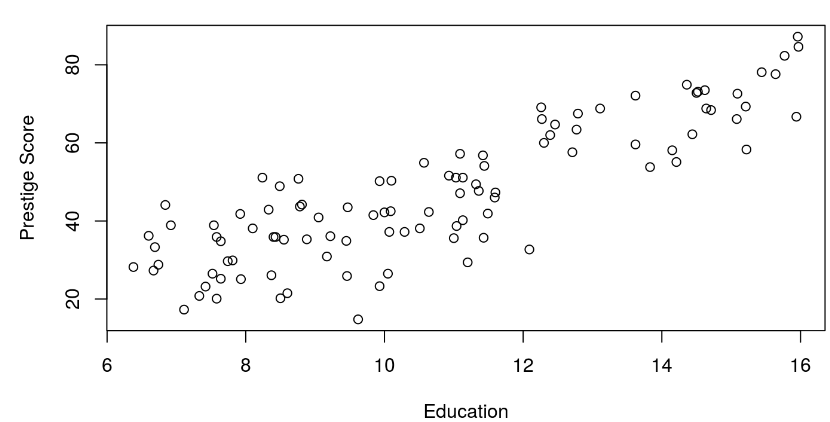


Years of education does not seem associated with % of women

```r
cor(Prestige$education, Prestige$women)
```

```
## [1] 0.06185286
```

```r
plot(Prestige$education,
     Prestige$prestige,
     xlab = 'Education',
     ylab = 'Prestige Score')
```



Education and prestige score are strongly associated!

```r
cor(Prestige$education, Prestige$prestige)
```

```
## [1] 0.8501769
```