# Week 3 – Discrete random variables

Juan Pablo Lewinger

# Last class

- Countable sample spaces (e.g. flipping a coin until first head, $\Omega = \{1, 2, \ldots\} = \mathbb{N}$)
- Conditional Probability: $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
- Law of total probability: $P(B) = P(B \mid A_1)P(A_1) + \ldots + P(B \mid A_n)P(A_n)$ for a (disjoint) partition $\Omega = A_1 \cup \ldots \cup A_n$
- Bayes theorem: $P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$
- Independence: $P(A \cap B) = P(A)P(B)$ (or more intuitively $P(A \mid B) = P(A)$

Q: is disjoint the same as independent?

- For $n > 2$ events independence requires satisfying $2^n$ equations

# Discrete random variables

A discrete random variable is a function $X : \Omega \longrightarrow \mathbb{R}$ that takes a finite or countable number of values $x_1, x_2, \ldots$

E.g. The number in the upper face of the rolled die, the sum of two dice

Notation:

- $\{\omega : X(\omega) = x\} = \{X = x\}$
- $P(\{X = x\}) = P(X = x)$

# Probability mass function (pmf)

The pmf of a discrete random variable taking values $x_1, x_2, \ldots$ is the function

$p : \mathbb{R} \longrightarrow [0, 1]$

$p(x) = P(X = x)$ (Sometimes also denoted $p_x(x)$ or $f_X(x)$)

If $X$ takes on the values $x_1, x_2, \ldots$ then:

- $p(x_i) > 0$, $\quad p(x_1) + p(x_2) + \ldots = 1$,
- $p(x) = 0$ for all other $x$.

E.g. Fair coin flip

$$X = \begin{cases} 1 & Heads \\ 0 & Tails \end{cases}$$

$p(1) = \frac{1}{2},\ p(0) = \frac{1}{2},\ p(0.5) = p(\pi) = 0$

# Cumulative distribution function

The cdf of a discrete random variable taking values $x_1, x_2, \ldots$ is the function

$F : \mathbb{R} \longrightarrow [0, 1]$ (sometimes denoted $F_X$)

$F(x) = P(X \leq x) \quad \forall x \in \mathbb{R}$

If $X$ takes on the values $x_1, x_2, \ldots$ then:

$F(x) = \sum_{x_i \leq x} p(x_i)$

E.g. Fair coin flip

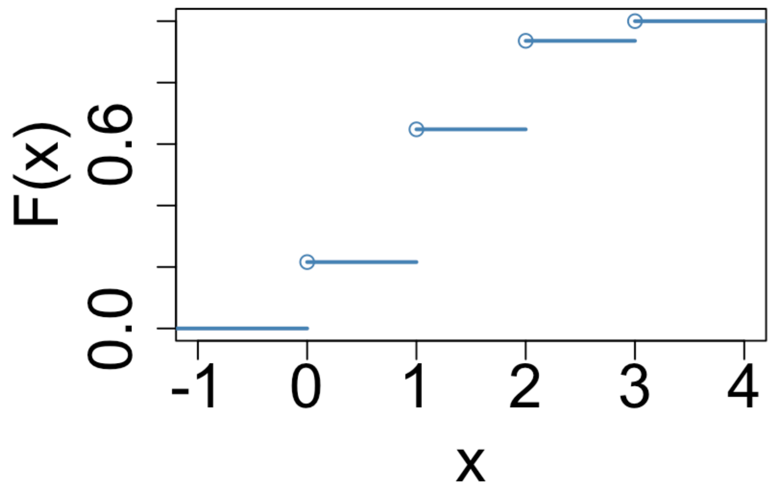
$$X = \begin{cases} 1 & Heads \\ 0 & Tails \end{cases}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2} & 0 \leq x < 1 \\ 1 & x \leq 1 \end{cases}$$

Both the pmf and the cdf completely characterize all the **probabilistic** information about a random variable (two random variables can have the same pmf and cdf and be different)

# Properties of the cumulative distribution function

- $F(x)$ is increasing: $x_1 \leq x_2 \implies F(x_1) \leq F(x_2)$
- $0 \leq F(x) \leq 1, \quad \lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to +\infty} F(x) = 1$
- $F$ is right continuous: $\lim_{\epsilon \downarrow 0} F(x + \epsilon) = F(x)$



# Bernoulli distribution

A random variable has a Bernoulli distribution

$f_X(1) = p, \quad (0 \leq p \leq 1)$

$f_X(0) = 1 - p$

$X \sim Bern(p)$ or $X \sim Bernoulli(p)$

Models experiments with only two possible outcomes

E.g. coin toss (H vs. T), die comes up six (yes vs. no)

A random variable with a Bernoulli distribution is called a Bernoulli trial

# Binomial distribution

$n$ independent Bernoulli trials (e.g. flipping a coin $n$ times) $X_1, \ldots, X_n \sim Bernoulli(p)$

$X = X_1 + \ldots + X_n$ counts the number of successes in $n$ trials

$X \sim Bin(n, p)$ or $X \sim Binomial(n, p)$

$f_X(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \; k = 0, 1, \ldots, n$

Models the number of successes in $n$ trials

For $n = 1$ the Binomial distribution is the Bernoulli distribution

# Binomial distribution

Example:

For example, suppose it is known that 5% of adults who take a certain medication experience negative side effects. What is the probability that more than $k$ patients in a random sample of 100 will experience negative side effects?

$$P(X > 1 \text{ patients experience side effects}) = ?$$

$$P(X > 5 \text{ patients experience side effects}) = ?$$

$$P(X > 15 \text{ patients experience side effects}) = ?$$

# Binomial distribution in R

pmf, cdf, and Random generation of a binomial random variable

```
dbinom(3, size=10, prob= 0.3)
```

```
## [1] 0.2668279
```

```
pbinom(3, size=10, prob= 0.3)
```

```
## [1] 0.6496107
```

```
rbinom(n=1, size=10, prob= 0.3)
```

```
## [1] 2
```

```
rbinom(n=3, size=10, prob= 0.3)
```

```
## [1] 0 1 3
```

# Binomial distribution

Example (continued):

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F_X(1)$$

```
1 - pbinom(1, size = 100, prob = 0.05)
```

```
## [1] 0.9629188
```

```
pbinom(1, size = 100, prob = 0.05, lower.tail = FALSE)
```

```
## [1] 0.9629188
```

```
pbinom(5, size = 100, prob = 0.05, lower.tail = FALSE)
```

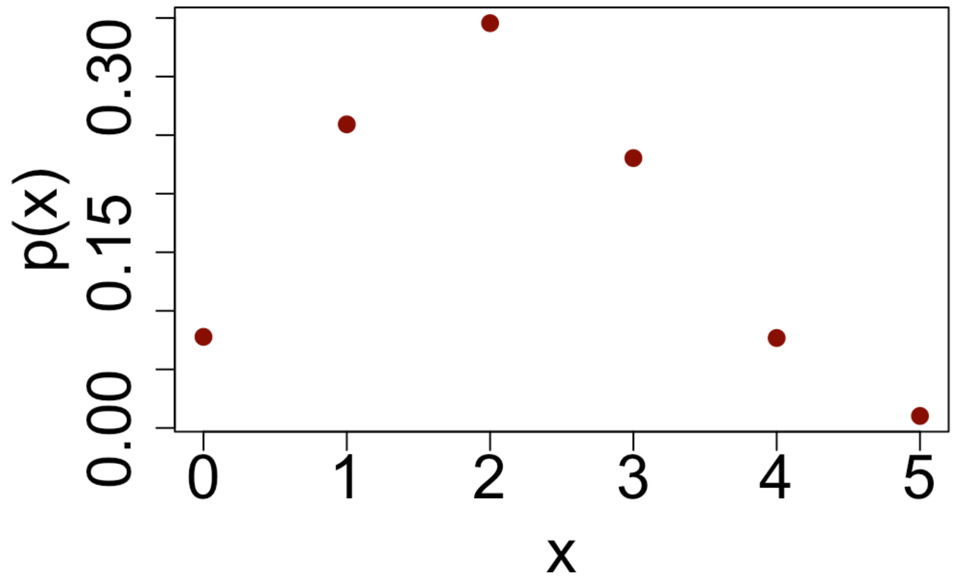
```
## [1] 0.3840009
```

```
pbinom(15, size = 100, prob = 0.05, lower.tail = FALSE)
```

```
## [1] 3.705408e-05
```

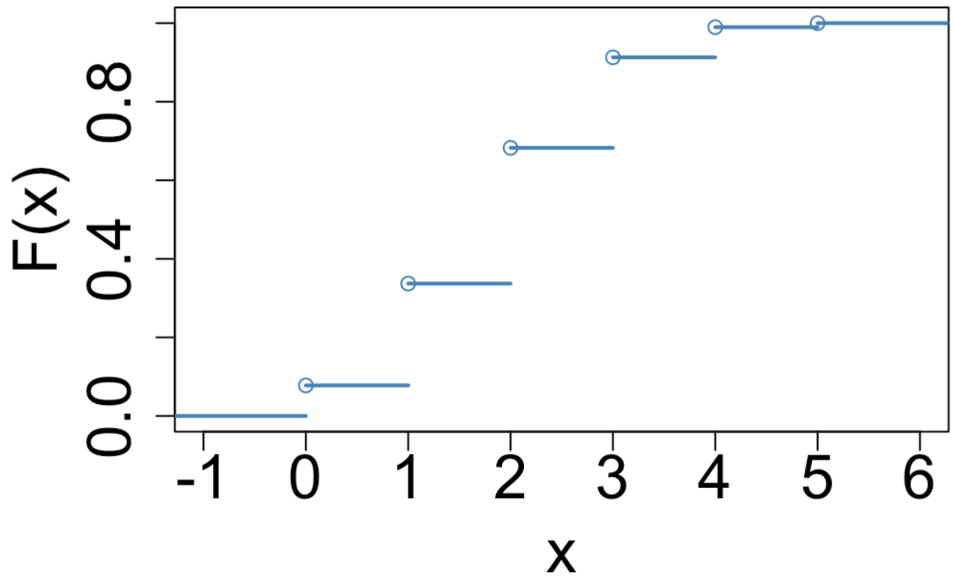
# Binomial distribution pmf

```
par(mar=c(6,8,5,1))

plot(0:5, dbinom(0:5, size=5, prob=0.4), col='red4', type='p', pch=16, cex=1.3, xlab='x', ylab='p(x)',
     cex.lab=2, cex.axis=2)
```



# Binomial distribution cdf

```
par(mar=c(6,8,5,1))

plot(stepfun(0:5, c(0, pbinom(0:5, size=5, prob=0.4))), pch = 1, lwd=2,
     col='steelblue', xlab='x', ylab='F(x)', cex.lab=2, cex.axis=2, main='', verticals = F)
```



# Simulating a binomial 3 different ways

Generating $n$ Bernoulli trials

```
Bernoulli_trials_1 = sample(0:1, 10, replace = TRUE, prob=c(0.3, 0.7))
Bernoulli_trials_1
```

```
##  [1] 1 1 0 1 1 0 1 0 1 0
```

```
sum(Bernoulli_trials_1)
```

```
## [1] 6
```

```
Bernoulli_trials_2 = rbinom(10, size=1, prob=0.3)
Bernoulli_trials_2
```

```
##  [1] 0 1 0 0 0 0 0 0 0 0
```

```
sum(Bernoulli_trials_2)
```

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

# Simulating a binomial 3 different ways

Directly sampling from the binomial

```
rbinom(1, size=10, prob=0.3)
```

```
## [1] 5
```

# Geometric distribution

A discrete random variable $X$ has a geometric distribution with parameter $p$, where $0 < p \leq 1$, if its probability mass function is given by:

$$p_X(k) = P(X = k) = (1 - p)^{k-1} p \qquad k = 1, 2, \ldots$$

$X \sim Geo(p)$ or $X \sim Geometric(p)$

Models the (discrete) waiting time until an event happens. E.g. number of trials till first heads

# Geometric distribution

Example:

You and a friend want to go to a concert, but there's only one ticket left. The salesperson decides to toss a coin until heads appears. In each toss heads appears with probability $p$, where $0 < p < 1$, independent of each of the previous tosses. If the number of tosses needed is odd, your friend is allowed to buy the ticket; otherwise you can buy it. Would you agree to this arrangement?

# Geometric memoriless property

$P(X > n + k \mid X > k) = P(X > n)$

The probability it'll take $n$ additional trials if the first $k$ are failures is the same as the probability it'll take $n$ trials at the begining of the experiment

# Geometric distribution in R

pmf, cdf, and Random generation of a binomial random variable

**Warning**: the definition of the geometric in R is the number of failures before the first success, i.e. $X - 1$

```
dgeom(x=5, prob = 0.1)
```

```
## [1] 0.059049
```

```
pgeom(10, prob= 0.1)
```

```
## [1] 0.6861894
```

```
rgeom(n=1, prob= 0.1)
```

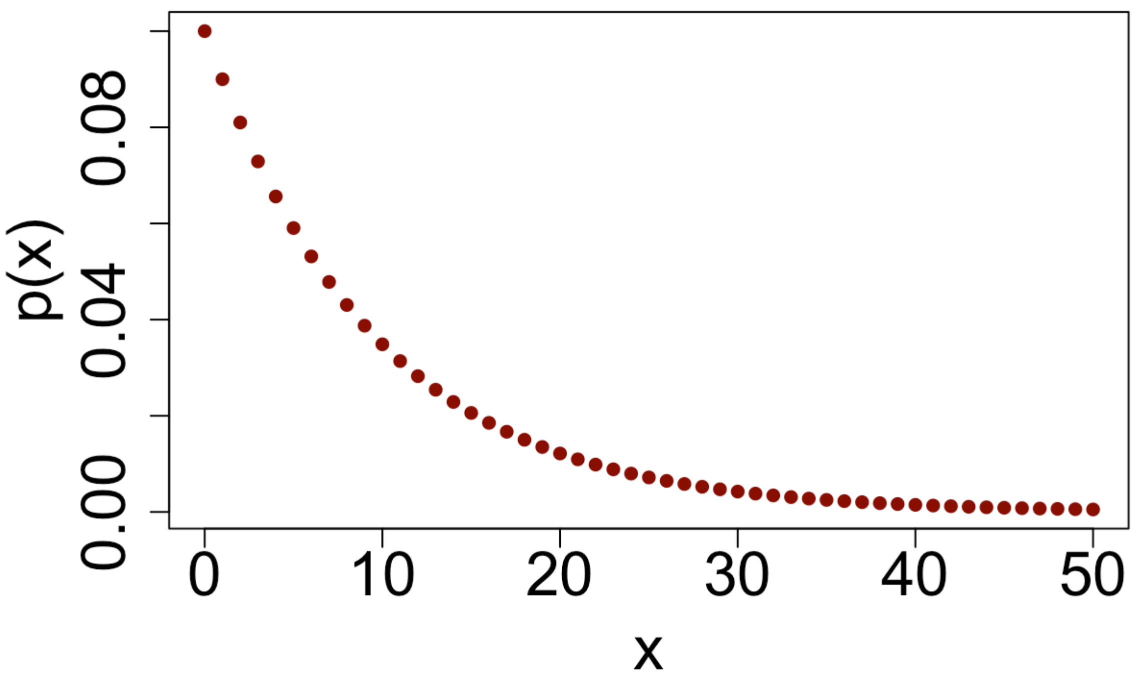
```
## [1] 0
```

```
rgeom(n=3, prob= 0.1)
```

```
## [1] 4 3 4
```

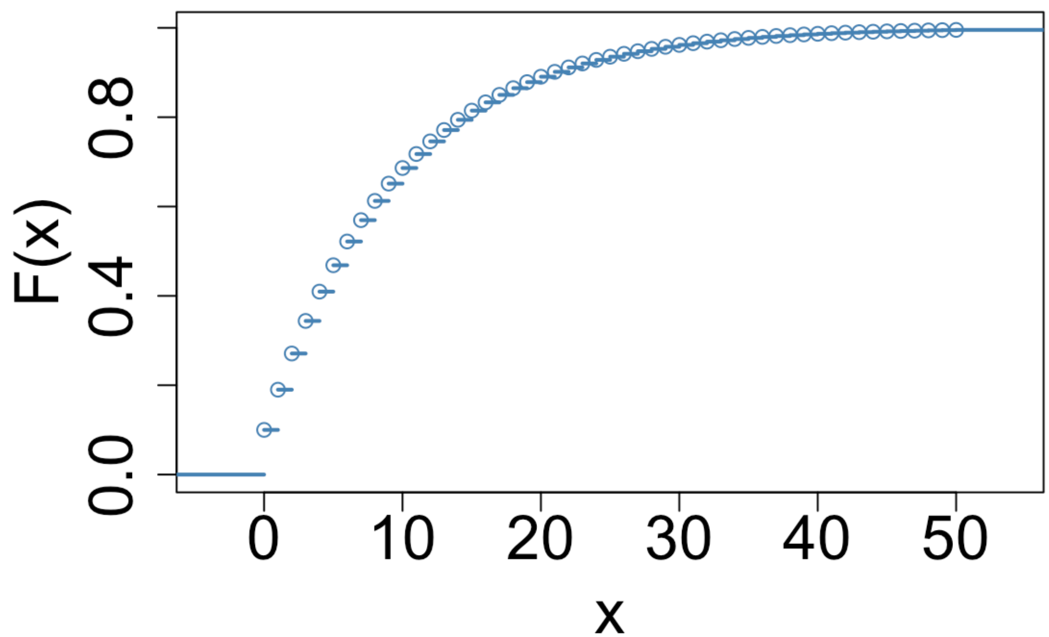
# Geometric distribution pmf

```
par(mar=c(6,8,5,1))

plot(0:50, dgeom(0:50, prob=0.1), col='red4', type='p', pch=16, cex=1, xlab='x', ylab='p(x)', cex.lab=2, cex.axis=2
```



# Geometric distribution cdf

```
par(mar=c(6,8,5,1))

plot(stepfun(0:50, c(0, pgeom(0:50, prob=0.1))), pch = 1, lwd=2, col='steelblue',
     xlab='x', ylab='F(x)', cex.lab=2, cex.axis=2, main='', verticals = F)
```



# Next week

- Read *IPS 4.0, 4.1.0-4.1.1* from ‘Introduction to probability, statistics, and random processes’ (do the exercises!)