Discrete Probability Spaces

C. Durso

Probability Spaces

- Common question in inference: Is there an effect in the data, or are they consistent with chance (null hypothesis)?
- Testing a null hypothesis requires calculation of the probability that data as extreme as the observed data occur under the null hypothesis.
- Probability distributions are essential to hypothesis testing.

Probability Spaces

- Some discrete probability distributions are in the standard Computer Science curriculum. Others are less familiar.
- Few, if any, continuous distributions are in the standard Computer Science curriculum.

Discrete Probability Spaces

- A discrete probability space (S, M, P) consists of:
  - The **sample space**: a finite or countable set, S, the set of possible outcomes
  - The **set of events**: a set M of subsets to S
  - The **probability function**: a function P from M to the real numbers in the interval $[0,1]$
- Example: Model the result of rolling a fair die as a discrete probability space: $S = \{1, 2, 3, 4, 5, 6\}$, M is the power set of S, and $P(A) = \frac{1}{6}|A|$.

Conditions on M

- The empty set and S itself must be elements of M. If $\{A_1, A_2\} \subseteq M$ then
  - $A_1 \cup A_2 \in M$
  - $\overline{A_1} \in M$
- Further, if $\{A_1, A_2, \dots A_n, \dots\} \subseteq M$, then $\cup_{i=1}^{\infty} A_i \in M$. Note that this implies that M is closed under countable union, complementation, and intersection.
- In practice, M will usually be the power set of S.

Conditions on P

- $P: M \to [0,1]$
- $P(\phi) = 0$
- $P(S) = 1$
- If $A, B \in M$ are disjoint, then
$$P(A) + P(B) = P(A \cup B)$$
- If $\{A_1, A_2, \ldots A_n, \ldots\} \subseteq M$ and the A_i's are pairwise disjoint, then
$$P(\cup_{i=1}^{\infty} A_i) = \Sigma_{i=1}^{\infty} P(A_i)$$

Density f

If, for every $x \in S$ we have that $\{x\} \in M$, we define the density f associated with (S, M, P).

Definition

The discrete probability density function $f: S \to [0,1]$ is defined by the relation $f(x) = P(\{x\})$.

Independence

Definition

Events A and B in a probability space (S, M, P) are **independent** if $P(A \cap B) = P(A)P(B)$.

Definition

Events $\{A_1, A_2, \ldots A_n\}$ are **mutually independent** if, for any subset

$$\{B_1, B_2, \ldots B_n\} \subseteq \{A_1, A_2, \ldots A_n\}, P\left(\cap_{i=1}^{k} B_i\right) = \Pi_{i=1}^{k} P(B_i)$$

Example

In the fair die example, the sets $\{1,2,3\}$ and $\{2,4\}$ are independent:

$$P(\{1,2,3\}) = \frac{1}{2}, P(\{2,4\}) = \frac{1}{3}, P(\{2\}) = \frac{1}{6} = \left(\frac{1}{2}\right)\left(\frac{1}{3}\right).$$

Discrete Random Variables

C. Durso

Discrete Random Variables

- For a discrete probability space (S, M, P) for which M is the power set of S, any function $X: S \longrightarrow \mathbb{R}$ is called a **discrete random variable**.
- Such a discrete random variable gives rise to a discrete probability space with sample space equal to the range of X and the set of events equal to the power set of the range of X.
- The probability function is determined by the density, $f(x) = P(\{s \in S | X(s) = x\})$.

Example: Bernoulli Trial

Definition

A Bernoulli Trial with probability of success p is a discrete probability space with

- $S = \{\text{success}, \text{failure}\}$
- M is the power set of S
- $f(\text{success}) = p, f(\text{failure}) = 1 - p$

Example: Multiple Bernoulli Trials

Define a discrete probability space to model the experiment of n repetitions of independent Bernoulli trials with probability of success p:

- S is the set of all sequences of n total successes and failures, $(a_1, a_2, \ldots a_n)$ where $a_i \in \{\text{success}, \text{failure}\}$ for all i.
- M is the power set of S.
- P is defined by the density f with $f\big((a_1, a_2, \ldots a_n)\big) = p^k(1-p)^{n-k}$ when there are k successes in $(a_1, a_2, \ldots a_n)$.

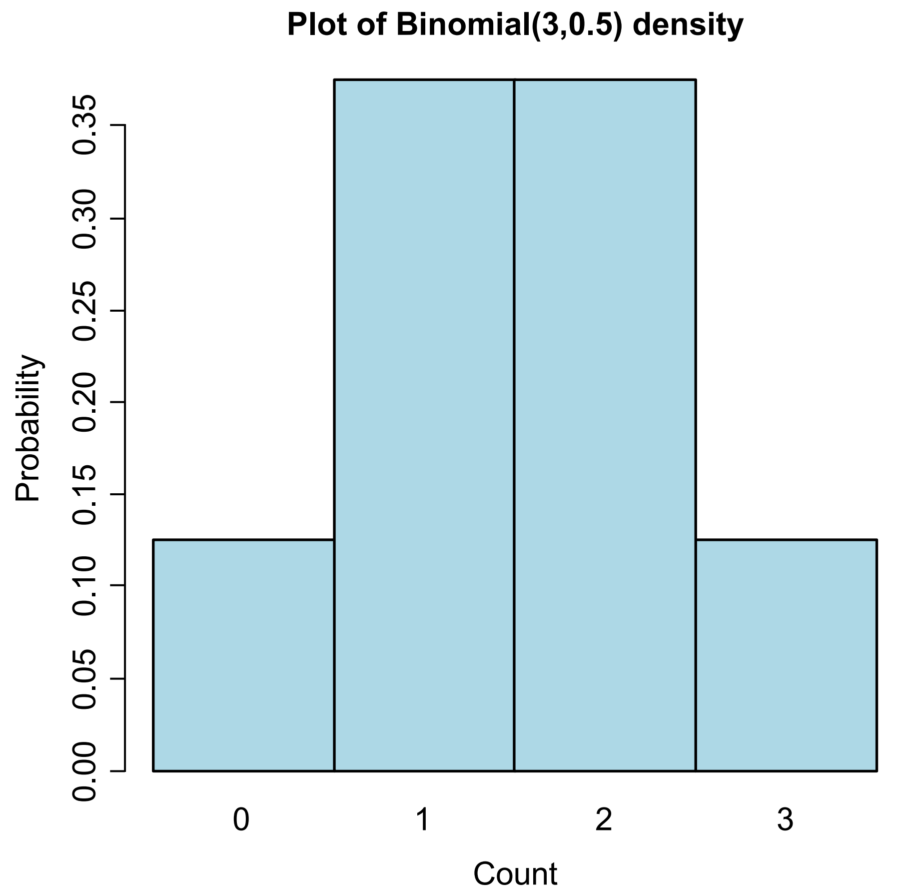
Example: Random Variable

- Given the previous discrete probability space modeling the sequence of successes and failures in n independent Bernoulli trials with probability of success p, define the random variable $X: S \longrightarrow \mathbb{R}$ by $X\big((a_1, a_2, \ldots a_n)\big)$ equals the count of successes in $(a_1, a_2, \ldots a_n)$.
- The probability of the event $\{s | X(s) = k\}$ for $k \in \{0,1,2, \ldots n\}$ is $\binom{n}{k} p^k (1-p)^{n-k}$
  - Every sequence with k successes has probability $p^k (1-p)^{n-k}$
  - There are $\binom{n}{k}$ sequences with k successes: $\binom{n}{k}$ counts possibilities for locations of successes

Example: Binomial Distribution

- This gives rise to a new probability space $(\tilde{S}, \widetilde{M}, \tilde{P})$.
  - The set of outcomes equal to $\{0,1,2, \ldots n\}$
  - The set of events equal to the power set of the set of outcomes
  - The density function of $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$
- This is known as the binomial distribution
- $\boldsymbol{Binomial(n, p)}$. This is a family of discrete distributions with two parameters. Think of the number of heads in n independent tosses of a possibly biased coin with p equal to the probability of heads.

Plot of a Binomial Distribution



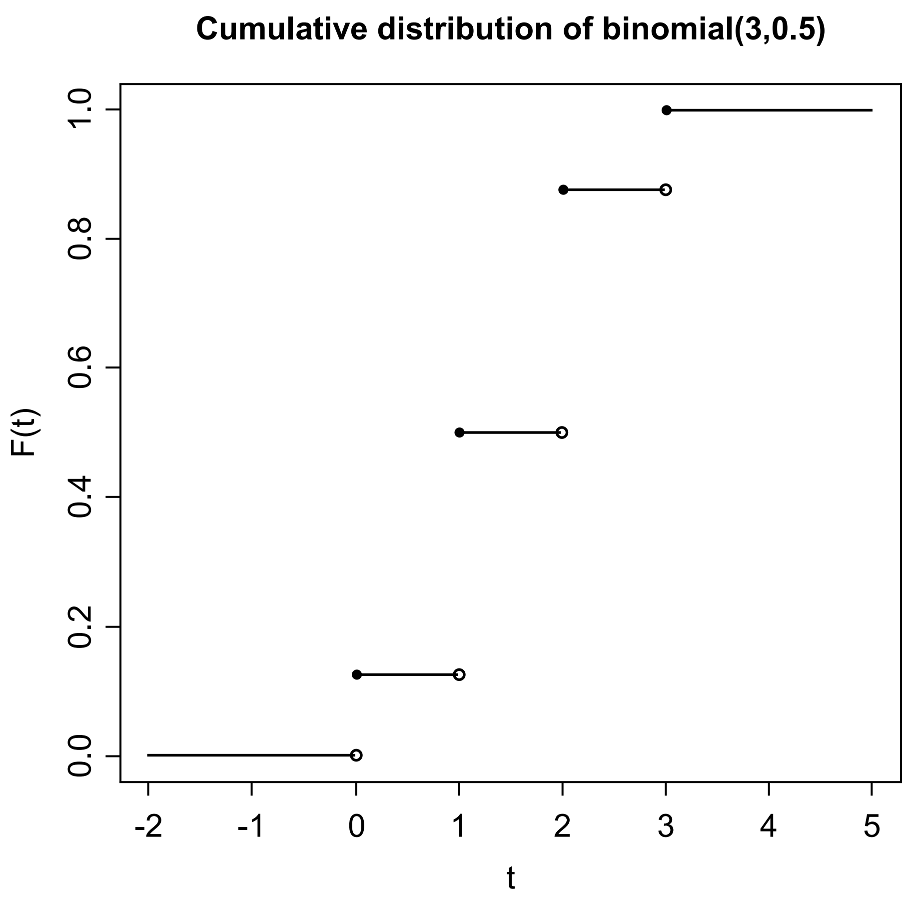
Cumulative Distribution

For a discrete probability space (S, M, P) with $S \subseteq \mathbb{R}$ and M equal to the power set of S, we define the cumulative distribution of (S, M, P).

Definition

The cumulative distribution of (S, M, P) is the function $F: \mathbb{R} \longrightarrow [0,1]$ by $F(t) = P(\{x \in S | x \leq t\})$.

Example of Cumulative Distribution



Definition of Continuous Probability Distributions

C. Durso

Continuous Probability Spaces

- A continuous probability space (S, M, P) consists of:
  - The **sample space**: a *measurable* set, $S \subseteq \mathbb{R}$, the set of possible outcomes
  - The **set of events**: a set M of subsets of $\mathbb{R}$, usually all *measurable* sets
  - The **probability function**: a function P from M to the real numbers in the interval $[0,1]$
- Example: Model the selection of a value in $[0,1]$ in which any value is equally likely. M includes all intervals in S, and $P([a, b]) = b - a$ for all a, b with $0 \leq a \leq b \leq 1$.

Conditions on M, Continuous

- The empty set and S itself must be elements of M. If $\{A_1, A_2\} \subseteq M$ then
  - $A_1 \cup A_2 \in M$
  - $\overline{A_1} \in M$
- Further, if $\{A_1, A_2, \ldots A_n, \ldots\} \subseteq M$ then $\cup_{i=1}^{\infty} A_i \in M$. Note that this implies that M is closed under countable union, complementation, and intersection.
- In practice, M will include all intervals in S. Typically, M is all Lebesgue measurable sets.

Conditions on P, Continuous

- For a continuous probability space, there exists a measurable function f such that, for any $A \in M$, the value of $P(A)$ equals $\int_A f(x)dx$.
  - For all x, $f(x) \geq 0$
  - $\int_{\mathbb{R}} f(x)dx = 1$
  - If $x \notin S$ then $f(x) = 0$

Example: Uniform Distribution on $[0,1]$

The uniform distribution on $[0,1]$ has:

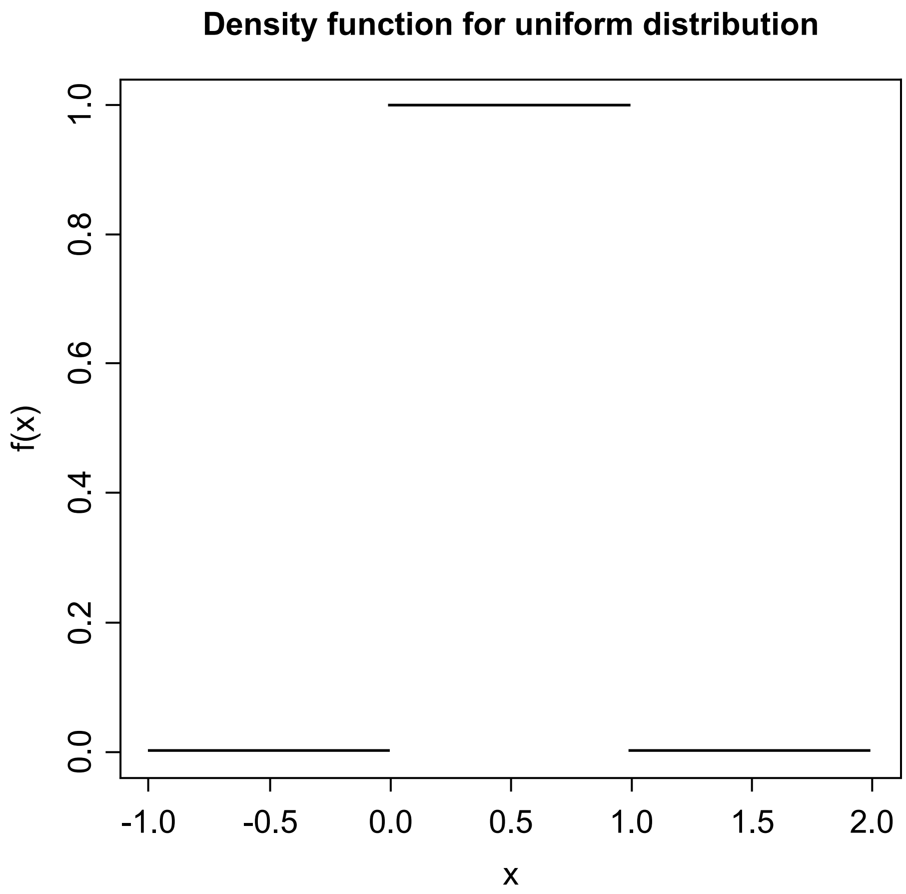
- $S = [0,1]$
- M is the set of Lebesgue measurable subsets of $\mathbb{R}$
- $f(x) = \begin{cases} 1 & x \in [0,1] \\ 0 & x \notin [0,1] \end{cases}$

Probability of an Interval

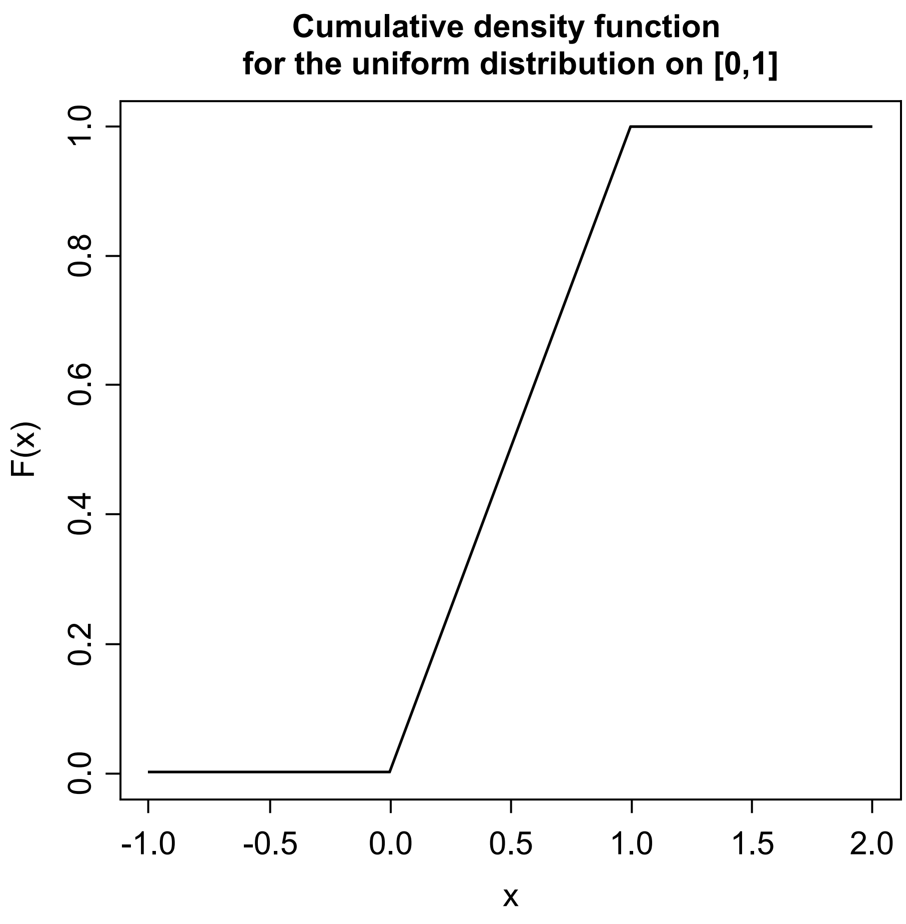
Under the uniform distribution,

- if $0 \le a \le b \le 1$, then the $P((a,b)) = \int_a^b 1dx = x\Big|_a^b = b - a$
- if $0 \le t \le 1$, then the $P((-\infty,t)) = \int_0^t 1dx = t$

Density Function



Cumulative Distribution



General Uniform Distribution

- More generally, the uniform distribution on $[a, b]$ has

$$f(x) = \begin{cases} \dfrac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}$$

- This gives us a parametrized family of probability distributions $\mathrm{Uniform}(a, b)$ with two parameters.

Definition of Normal Distributions

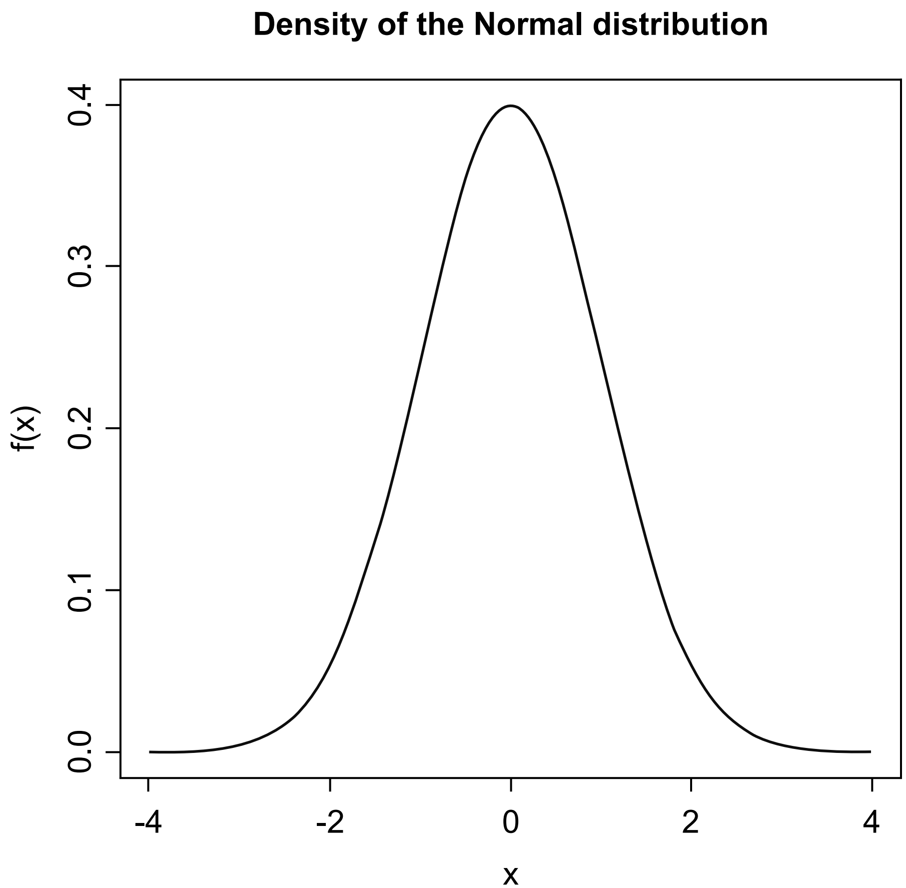
C. Durso

Standard Normal Distribution

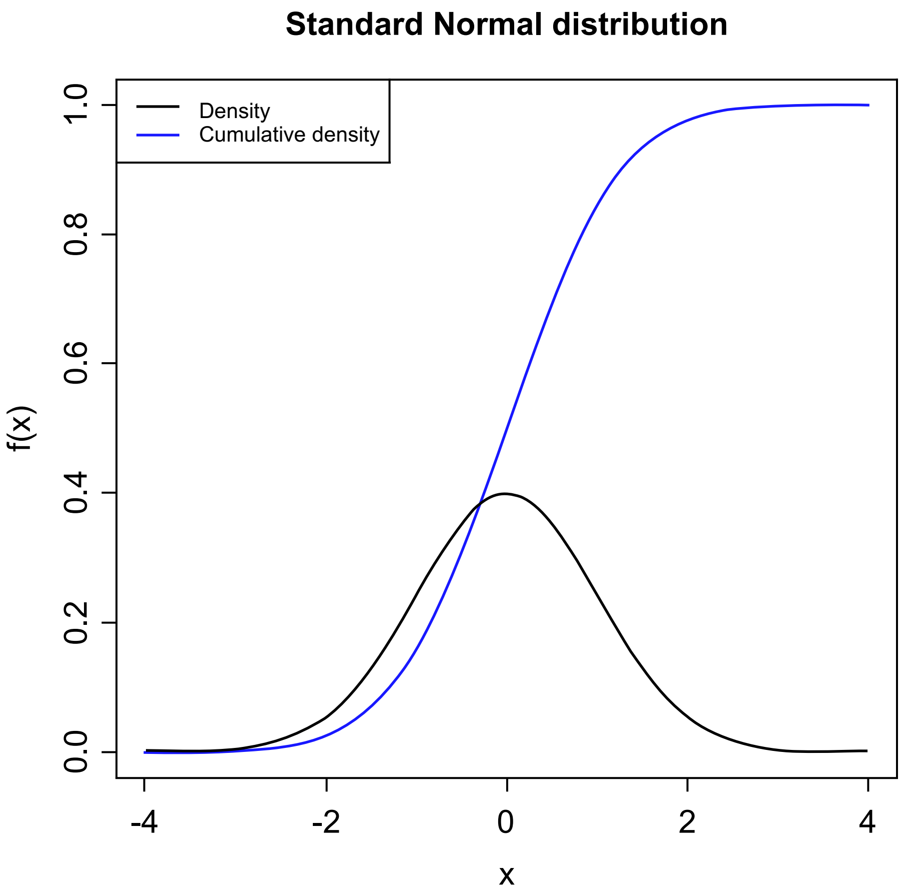
Standard Normal distribution, continuous distribution $\mathrm{Normal}(0,1)$:

- $S = \mathbb{R}$,
- $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$
  - Why does this have to be multiplied by $\frac{1}{\sqrt{2\pi}}$?
  - This keeps the integral over $\mathbb{R}$ equal to 1.

Standard Normal Density



Cumulative Density of Standard Normal



$\text{Normal}(\mu, \sigma^2)$

Set center and spread by specifying μ and σ^2

- Sample space $S = \mathbb{R}$,
- $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
  - Why does this have to be multiplied by $\frac{1}{\sqrt{2\pi\sigma^2}}$?

Verify Integral

Use the change of variable $u = \frac{x-\mu}{\sigma}, du = \frac{1}{\sigma}dx$