# 5: Continuous Random Variables

Spring 2023

Matthew Blackwell

Gov 2002 (Harvard)

# Where are we? Where are we going?

- Last few weeks: discrete random variables.
  - How to characterize uncertainty about data that takes on discrete values.
- Learned how to define distributions (p.m.f., c.d.f.) and how to summarize.
- Now: define the same ideas for r.v.s that can take on any real value.

# 1/ Continuous distributions

# Continuous r.v.s

- Discrete r.v.: specify $\mathbb{P}(X = x)$ for all possible values $\rightsquigarrow$ p.m.f.
- What if $X$ can take any value on any real value?
- Can we just specify $\mathbb{P}(X = x)$ for all $x$?
- No! Proof by counterexample:
  - Suppose $\mathbb{P}(X = x) = \varepsilon$ for $x \in (0, 1)$ where $\varepsilon$ is a very small number.
  - What's the probability of being between 0 and 1?
  - There are an infinite number of real numbers between 0 and 1:

$$0.232879873\ldots \qquad 0.57263048743\ldots \qquad 0.9823612984\ldots$$

  - Each one has probability $\varepsilon \rightsquigarrow \mathbb{P}(X \in (0, 1)) = \infty \times \varepsilon = \infty$
- But $\mathbb{P}(X \in (0, 1))$ must be less than 1! $\rightsquigarrow \mathbb{P}(X = x)$ must be 0.

Thought experiment: draw a random real value between 0 and 10. What's the probability that we draw a value that is exact equal to $\pi$?

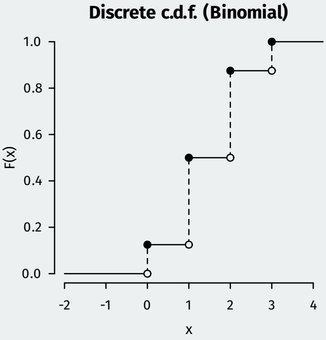
3.1415926535 8979323846 2643383279 5028841971 6939937510 5820974944 5923078164
0628620899 8628034825 3421170679 8214808651 3282306647 0938446095 5058223172
5359408128 4811174502 8410270193 8521105559 6446229489 5493038196 4428810975
6659334461 2847564823 3786783165 2712019091 4564856692 3460348610 4543266482
1339360726 0249141273 7245870066 0631558817 4881520920 9628292540 9171536436
7892590360 0113305305 4882046652 1384146951 9415116094 3305727036 5759591953
0921861173 8193261179 3105118548 0744623799 6274956735 1885752724 8912279381
8301194912 9833673362 4406566430 8602139494 6395224737 1907021798 6094370277
0539217176 2931767523 8467481846 7669405132 0005681271 4526356082 7785771342
7577896091 7363717872 1468440901 2249534301 4654958537 1050792279 6892589235
4201995611 2129021960 8640344181 5981362977 4771309960 5187072113 4999999837
2978049951 0597317328 1609631859 5024459455 3469083026 4252230825 3344685035
2619311881 7101000313 7838752886 5875332083 8142061717 7669147303 5982534904
2875546873 1159562863 8823537875 9375195778 1857780532 1712268066 1300192787
6611195909 2164201989 3809525720 1065485863 2788659361 5338182796 8230301952
0353018529 6899577362 2599413891 2497217752 8347913151 5574857242 4541506959
5082953311 6861727855 8890750983 8175463746 4939319255 0604009277 0167113900

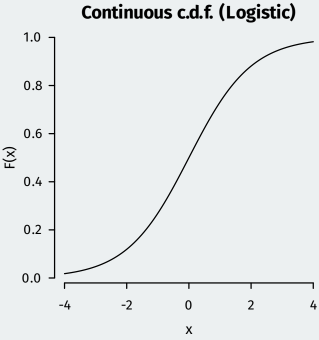
# Probability density functions

## Definition

A r.v., $X$, is **continuous** if its c.d.f. $F_X(x) = \mathbb{P}(X \leq x)$ is a continuous function.

- Essentially: the c.d.f. of a continuous r.v. has no jumps:



# Why "continuous"?

- How does a continuous c.d.f. connect to $\mathbb{P}(X = x)$? Note:

$$\mathbb{P}(X = x) \leq \mathbb{P}(x - \epsilon < X \leq x) = F_X(x) - F_X(x - \epsilon)$$

- But whe the c.d.f. is continuous we know that

$$\mathbb{P}(X = x) \leq \lim_{\epsilon \to 0} F(x) - F(x - \epsilon) = 0$$

- Continuous c.d.f.s imply the "point probabilities" are 0. What to do?
- With discrete, we summed up the p.m.f. to get the c.d.f.

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{j : x_j \leq x} p_X(x_j)$$

- For continuous r.v.s, we'll replace the sum with an integral!

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t) dt$$

# Probability density function

**Definition**

The **probability density function** of a continuous r.v. $X$ $f_X(x)$ is the function that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \quad \text{for all } x.$$

- By the fund. theorem of calculus p.d.f. is the derivative of the c.d.f.:

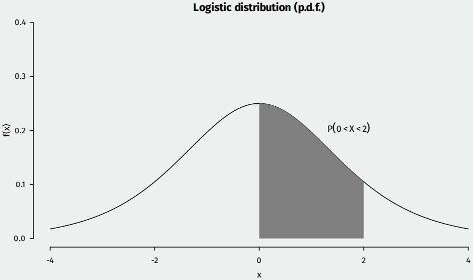
$$\frac{d}{dx} F_X(x) = f_X(x)$$

- Interval probabilities:

$$\mathbb{P}(a < X < b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F(b) - F(a) = \int_a^b f_X(x)dx$$

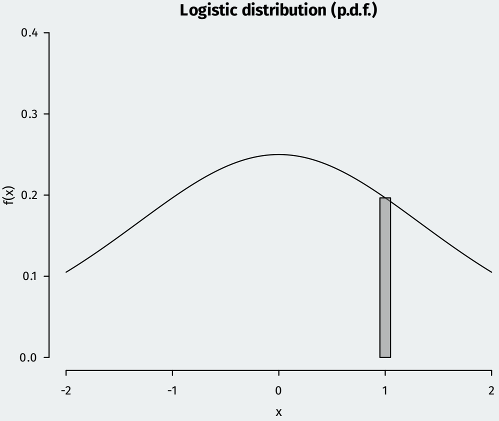
- With continuous we don't have to worry about $<$ vs $\leq$.
    - $\mathbb{P}(a < X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b)$.

# The p.d.f.




- $\rightsquigarrow$ the probability of a region is the area under the p.d.f. for that region.
  - Support of $X$ is all values such that $f_X(x) > 0$.
- Properties of a valid p.d.f.:
  - Nonnegative: $f_X(x) > 0$
  - Integrates to 1: $\int_{-\infty}^{\infty} f_X(x)dx = 1$
- **Important:** $f_X(x)$ can be bigger than 1!

# p.d.f. intuition




- Intuition of a density:

$$f(x_0)\varepsilon \approx \mathbb{P}(X \in (x_0 - \varepsilon/2, x_0 + \varepsilon/2))$$

# Continuous uniform distribution

- Simple and really important continuous distribution: **uniform**.
  - Intuitively, every equal-sized interval has the same probability.
  - How can figure out the p.d.f. for such a distribution?

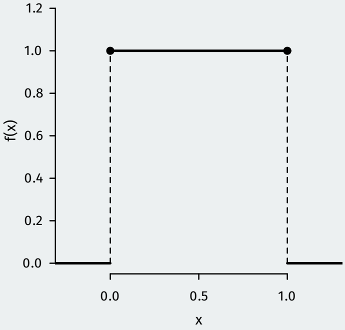
**Definition**

A continuous r.v. $U$ has a **Uniform distribution** on the interval $(a, b)$ if its p.d.f. is

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

- If $(c, d)$ is a subinterval of $(a, b)$ then $\mathbb{P}(U \in (c, d))$ is proportional to $c - d$
- Distribution of $U$ conditional on being in $(c, d)$ is $\text{Unif}(c, d)$.

# Uniform pdf and cdf




- **Location-scale transformation**: Let $U \sim \text{Unif}(a, b)$. Then $\widetilde{U} = cU + d$ is $\text{Unif}(ca + d, cb + d)$
  - Linear transformations of uniforms preserve the uniform distribution.

# **2/** Expectation for continuous r.v.s

# Expectation for a continuous r.v.

- Expectation of a continuous r.v.:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

  - Unifying notation you may see: $\mathbb{E}[X] = \int_{-\infty}^{\infty} x dF(x)$
  - Expectation of a uniform (0,1): $\mathbb{E}[U] = (a+b)/2$

- LOTUS with continuous r.v.s: $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$

- Variance of a continuous r.v.s:

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 dx$$

- Linearity and other properties of $\mathbb{E}[]$ and $\mathbb{V}[]$ still hold!
  - In particular, we still have $\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

# Expectation of random circle areas

- Let $R \sim \text{Unif}(0,1)$ and $A$ be the area of the circle with radius $R$.
- What are $\mathbb{E}[A]$ and $\mathbb{V}[A]$?
- For expectation, use LOTUS!

$$\begin{aligned}\mathbb{E}[A] = \mathbb{E}[\pi R^2] &= \int_0^1 \pi r^2 dr \\ &= (\pi/3) r^3 \Big|_0^1 \\ &= (\pi/3)\cdot 1^3 - (\pi/3)\cdot 0^3 = (\pi/3)\end{aligned}$$

- For variance, use $\mathbb{V}[A] = \mathbb{E}[A^2] - (\mathbb{E}[A])^2$:

$$\begin{aligned}\mathbb{E}[A^2] = \mathbb{E}[\pi^2 R^4] &= \int_0^1 \pi^2 r^4 dr = (\pi^2/5) r^5 \Big|_0^1 \\ &= (\pi^2/5)\cdot 1^5 - (\pi^2/5)\cdot 0^5 = (\pi^2/5)\end{aligned}$$

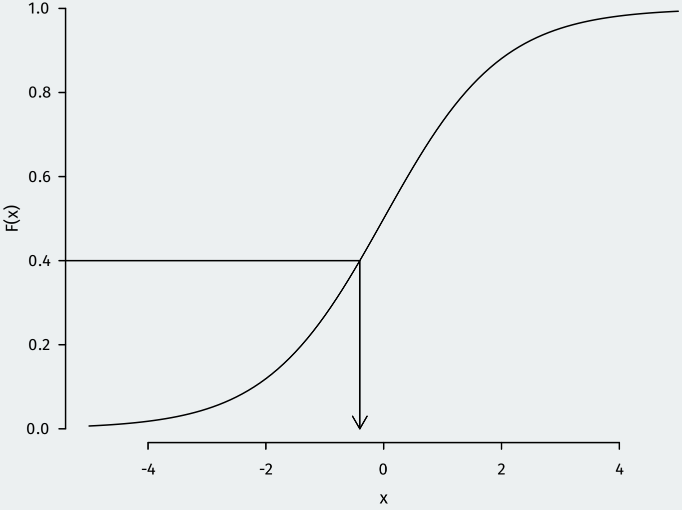
- $\rightsquigarrow \mathbb{V}[A] = 4\pi^2/45$. **Challenge:** find the c.d.f. and p.d.f. of $A$

# **3/** Universality of the uniform

# Quantile function

- Inverse of the c.d.f. $F^{-1}$ is called the **quantile function**
  - $F^{-1}(\alpha)$ is the value of $X$ such that $\mathbb{P}(X \leq x) = \alpha$
  - Takes probabilities as arguments!
  - $F^{-1}(0.5)$ is the median, $F^{-1}(0.25)$ is the lower quartile, etc
- Intuition: exactly the same as percentiles on exams.
- You've probably used them before: confidence interval critical values.

# Quantile functions



# Universality of the Uniform

- The Uniform distribution has a deep connection to all continuous r.v.s

1. Let $U \sim \text{Unif}(0, 1)$ and $X = F^{-1}(U)$, then $X$ is an r.v. with c.d.f. $F$.
2. If $X$ is an r.v. with c.d.f. $F$, then $F(X) \sim \text{Unif}(0, 1)$.

- **Careful**: $F(X)$ means plug the random variable into the c.d.f. as a function.
    - Not $F(X) \neq \mathbb{P}(X \leq X)$.

# 4/ Normal distribution

# Standard normal distribution

**Definition**

A continuous r.v. $Z$ follows a **standard normal distribution** if its p.d.f. $\varphi$ is given as

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad -\infty < z < \infty,$$

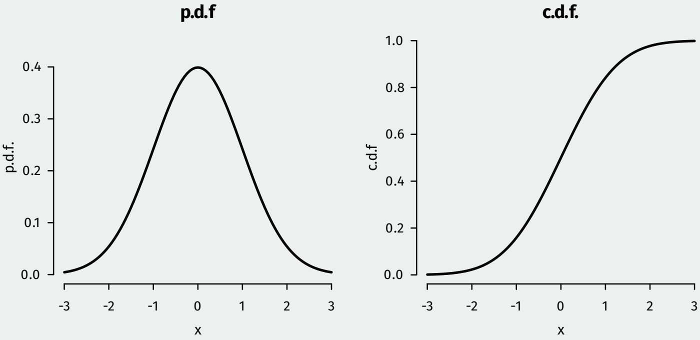
and we write this $Z \sim \mathcal{N}(0, 1)$

- Not immediately obvious, but tricky calculus will show $\int_{-\infty}^{\infty} \varphi(z) = 1$.
- Normal c.d.f. has no closed form solution, so written as:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

- Standard normal is mean zero, variance 1: $\mathbb{E}[Z] = 0, \mathbb{V}[Z] = 1$.

# The normal distribution




- Deeply symmetric:
  - p.d.f. is symmetric: $\varphi(z) = \varphi(-z)$
  - Tail areas are symmetric $\Phi(z) = 1 - \Phi(-z)$
  - $Z$ and $-Z$ are both $\mathcal{N}(0, 1)$

# General normal distribution

**Defintion**

If $Z \sim \mathcal{N}(0, 1)$ then

$$X = \mu + \sigma Z$$

follows the normal distribution with mean $\mu$ and variance $\sigma^2$, written $X \sim \mathcal{N}(\mu, \sigma^2)$.

- We can move back to a standard normal through **standardization**:

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

- c.d.f.: $\Phi((x - \mu)/\sigma)$
- p.d.f.:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

# Properties of normals and sums

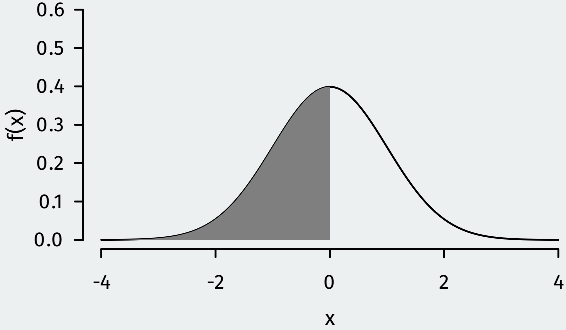
- If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and $X_1 \perp\!\!\!\perp X_2$,

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

- **Cramer's theorem:** if $X_1 \perp\!\!\!\perp X_2$ and $X_1 + X_2$ is normal, then $X_1$ and $X_2$ are normal.

# Using pnorm

- `pnorm( )` evaluates the c.d.f. of the normal:

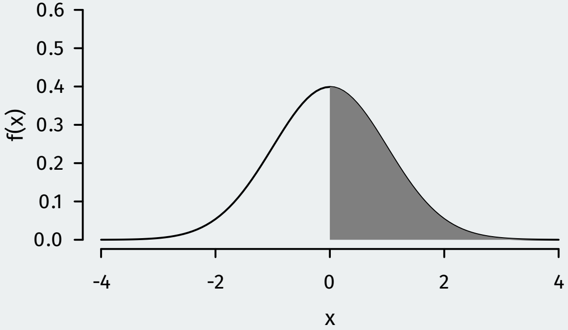



```r
pnorm(q = 0, mean = 0, sd = 1)
```

```
## [1] 0.5
```

# Using pnorm

- `pnorm( )` evaluates the c.d.f. of the normal:

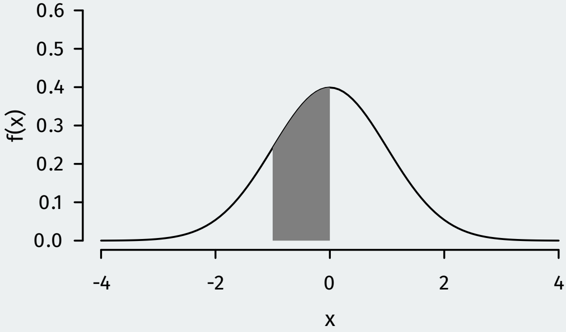



```r
pnorm(q = 0, mean = 0, sd = 1, lower.tail = FALSE)
```

```
## [1] 0.5
```

# Using pnorm

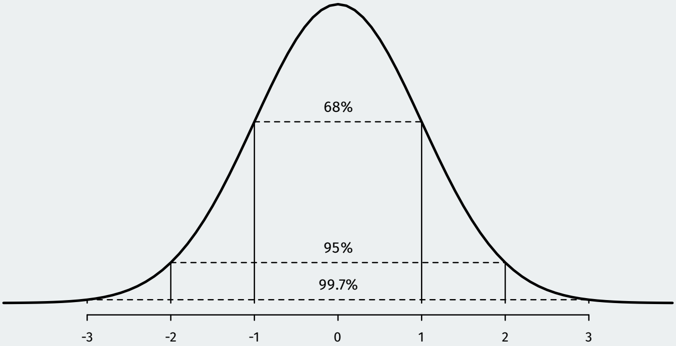
- `pnorm( )` evaluates the c.d.f. of the normal:




```r
pnorm(q = 0, mean = 0, sd = 1) - pnorm(q = -1, mean = 0, sd = 1)
```

```
## [1] 0.341
```

# Empirical Rule for the Normal Distribution




- If $Z \sim \mathcal{N}(0, 1)$, then the following are roughly true:
  - Roughly 68% of the distribution of $Z$ is between -1 and 1.
  - Roughly 95% of the distribution of $Z$ is between -2 and 2.
  - Roughly 99.7% of the distribution of $Z$ is between -3 and 3.

# Chi-square distribution

**Definition**

Let $V = Z_1^2 + \cdots + Z_n^2$ where $Z_1, Z_2, \dots, Z_n$ are i.i.d. $\mathcal{N}(0,1)$. Then $V$ follows the **Chi-square distribution** with $n$ degrees of freedom, written $V \sim \chi_n^2$

- Why do we care? **Sample variance** of normal r.v.s $X_1, \dots, X_n$ i.i.d. $N(\mu, \sigma^2)$:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

- Furthermore, $\overline{X}_n$ is independent of $s^2/\sigma^2$.

# Student t distribution

**Definition**

If $Z \sim \mathcal{N}(0,1)$ and $V \sim \chi^2_n$ with $Z \perp\!\!\!\perp V$, then

$$T = \frac{Z}{\sqrt{V/n}},$$

follows the **student-t distribution** with $n$ degrees of freedom, written $T \sim t_n$.

- Important result for the **normal model**: if $X_1, \ldots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$:

$$T = \frac{\overline{X}_n - \mu}{\sqrt{s^2/n}} \sim t_{n-1}$$

- Properties of the $t$ distribution:
  - Symmetric and mean-zero like the standard normal.
  - Fatter tails than the normal.
  - Converges to $\mathcal{N}(0,1)$ as $n \to \infty$

# Appendix

# Symmetry of iid continuous r.v.s

**Proposition**

Let $X_1, \dots, X_n$ be i.i.d. from a continuous distribution. Then,

$$\mathbb{P}(X_{a_1} < X_{a_2} < \cdots < X_{a_n}) = \frac{1}{n!}$$

for any permutation $a_1, a_2, \dots, a_n$ of $1, 2, \dots, n$.

- All orderings of continuous i.i.d. r.v.s are equally likely.
- Doesn't necessarily hold for discrete r.v.s