General random variables

James Scott (UT-Austin)

Reference: Bertsekas Chapters 3.1-3.3, 2.3, 3.6

Outline

- The CDF
- Continuous random variables and PDFs
- Moments for continuous random variables
- The inverse CDF
- The normal distribution
- New random variables from old ones

The CDF

Up to now we've only been dealing with discrete random variables that are characterized by a PMF.

But what about a random variable like:

- $X =$ Apple's stock price tomorrow?
- $X =$ speed of the next pitch thrown by Justin Verlander?
- $X =$ blood pressure of a randomly sampled participant in a clinical trial of a new drug?

These outcomes cannot naturally be restricted to a finite or countable set, and they don't have PMFs. To describe these random variables, we need some more general concepts.

The CDF

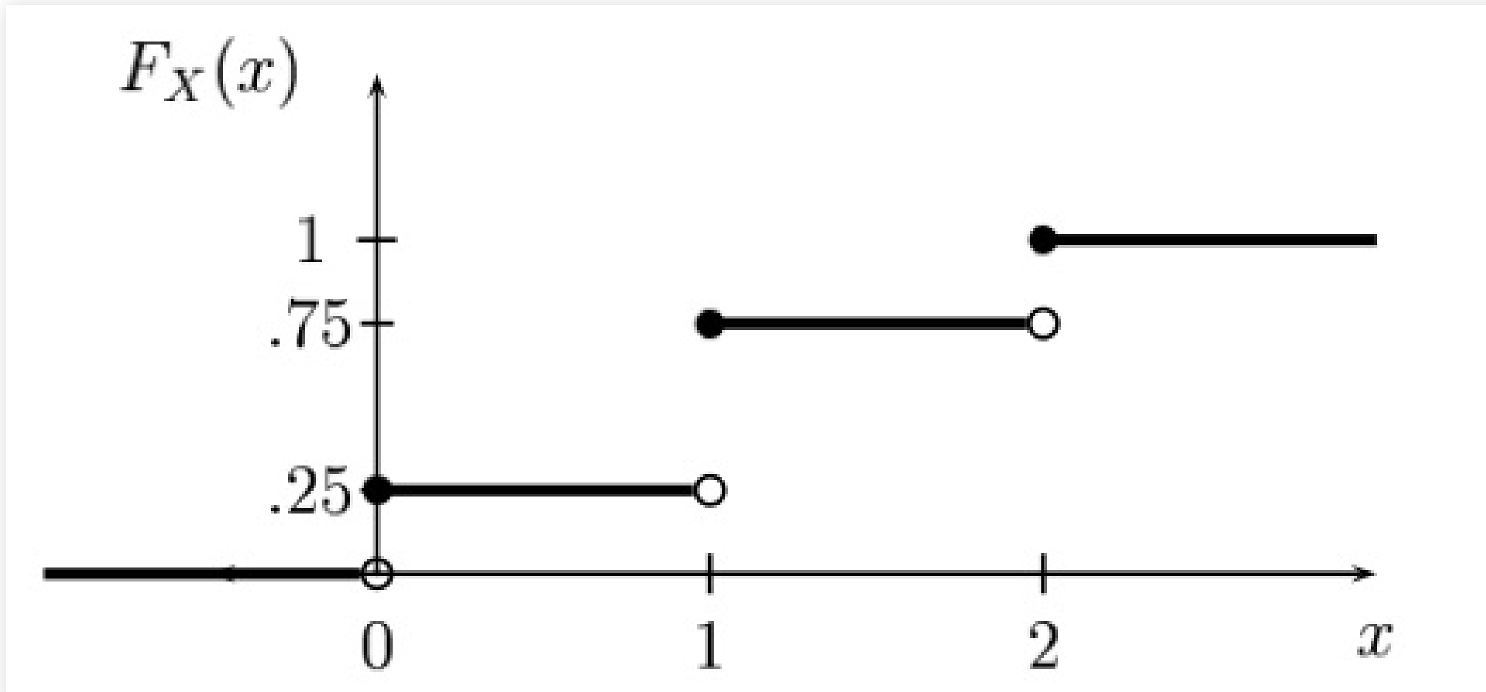
The cumulative distribution function, or CDF, is defined as:

$$F_X(x) = P(X \leq x)$$

Facts:

- *All* random variables have a CDF.
- The CDF completely characterizes the random variable: if X and Y have the same CDF, then for all sets S, $P(X \in S) = P(Y \in S)$.
- If this holds, we say that X and Y are *equal in distribution.* This doesn't mean they're identical! It just means that all probability statements about X and Y will be identical.

CDF: toy example



The CDF for Binomial(N=2, p=0.5). (Let's write this on the board.) The jumps correspond to the points where the PMF has positive probability. What is $F(1)$? What is $F(0.6)$? What is $F(17)$?

Properties of CDFs

All CDFs $F(x)$ satisfy the following properties:

1. F is non-decreasing: if $x_1 < x_2$, then $F(x_1) \leq F(x_2)$.
2. F is bottoms out at 0 and tops out at 1:
    - $\lim_{x \to -\infty} F(x) = 0$
    - $\lim_{x \to \infty} F(x) = 1$
3. F is right-continuous, i.e.

$$F(x) = \lim_{y \downarrow x} F(y)$$

Note: $\lim_{y \downarrow x}$ means "limit as y approaches x from above."

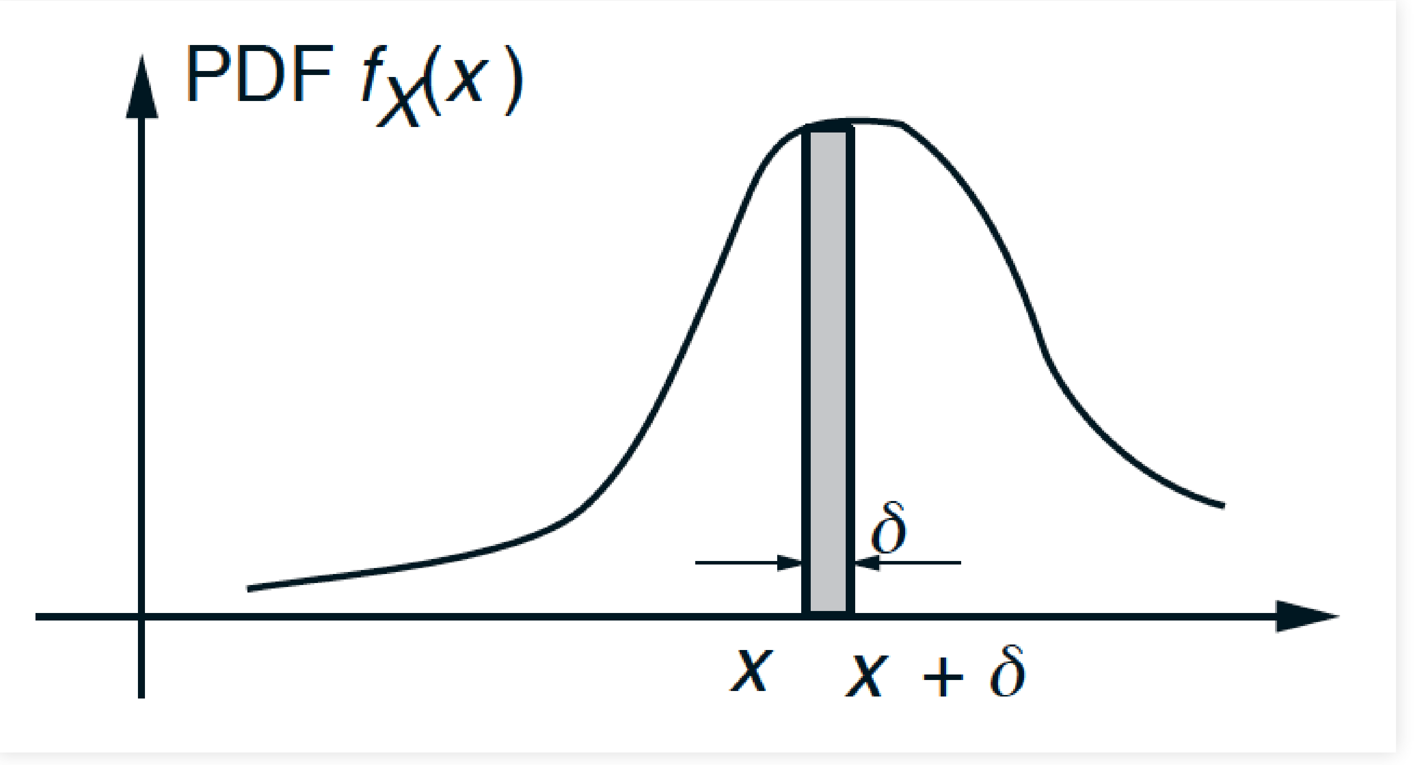
Continuous random variables

Intuitively, a continuous random variable is one that has no "jumps" in its CDF. More formally, we say that X is a continuous random variable if there exists a function f such that:

1. $f(x) \geq 0$
2. $\int_{\mathcal{R}} f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1$
3. For every interval $S = (a, b)$,

$$P(X \in S) = P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

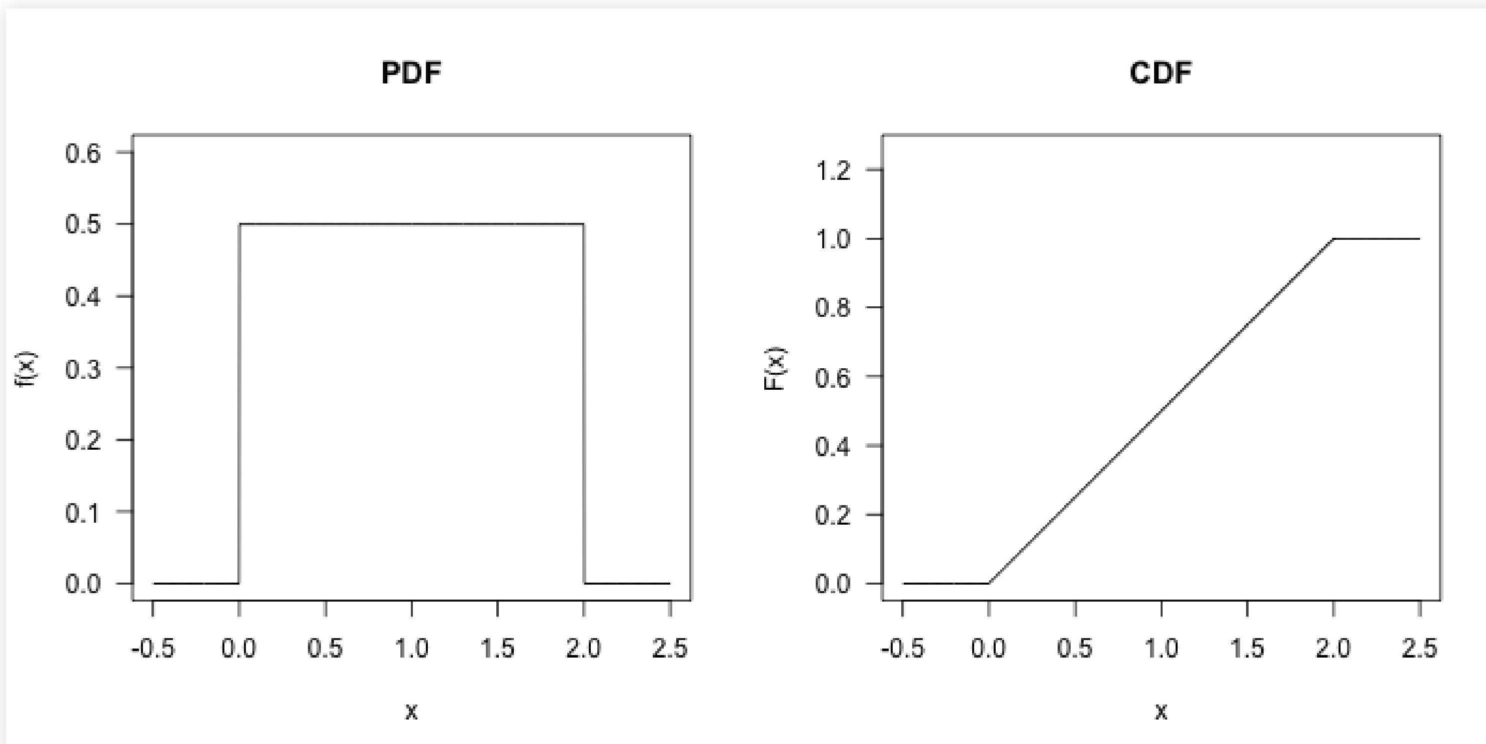
Continuous random variables



If δ is small, then $P(x < X < x + \delta) \approx f(x) \cdot \delta$. The PDF can be interpreted as "probability per unit length" (like density in physics).

Example 1: uniform distribution

Suppose that X is a random variable with PDF $f_X(x) = 1/2$ for $0 \leq x \leq 2$ (and $f(x) = 0$ otherwise). We write this as $X \sim$ Uniform(0, 2).



Example 1: uniform distribution

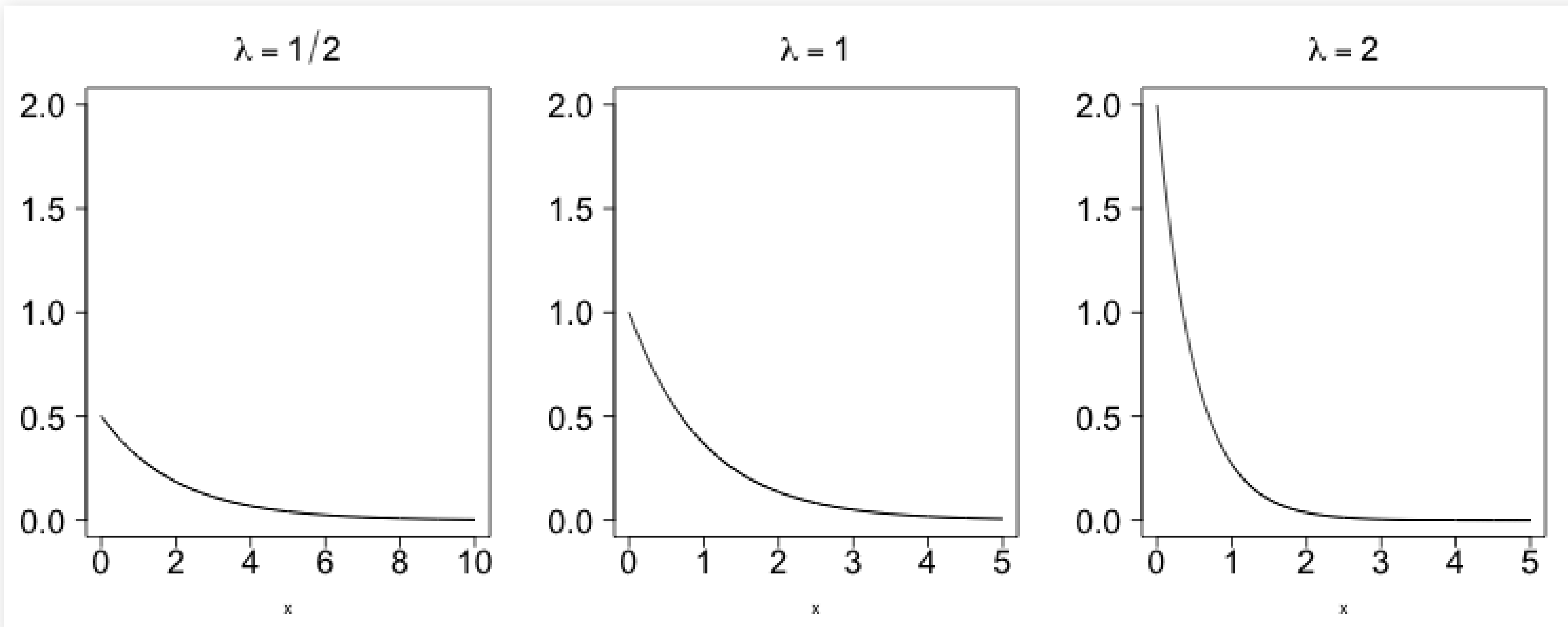
More generally, we say that X has a (continuous) uniform distribution on (a, b) if its PDF tales the form:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

We write this as $X \sim$ Uniform(a,b). Note: this is different than the *discrete* uniform distribution, which places probability $1/N$ on N discrete points.

Example 2: Exponential distribution

Suppose that X is a random variable with PDF $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$ (and $f(x) = 0$ otherwise).



Example 2: Exponential distribution

The exponential distribution also has a wide range of applications. For example:

- industry: the lifetime of manufactured components.
- science: photon emissions
- queueing: inter-arrival times (e.g. between customers in a store, hits on a website, etc.)
- transportation: wait times for ride-share pickups

The PDF/CDF relationship

Note that, by the definition of the CDF and PDF, we have the following relationship for a continuous random variable:

$$F_X(x) = P(X \in (-\infty, x)) = \int_{-\infty}^{x} f_X(x)dx$$

Remember the Fundamental Theorem of Calculus! This relationship says that the PDF is the derivative of the CDF:

$$f(x) = F'(x)$$

at all points where $F(x)$ is differentiable.

PDF/CDF: Example 1

Recall that if $X \sim$ Uniform(a,b), then the PDF is

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

The corresponding CDF can then be computed as:

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

PDF/CDF: Example 2

We can also go the other direction. For example:

- Suppose you give the same test to a group of 10 high-school students. Let X_i be the score for student i, and assume that the test scores are independent across students.
- Suppose that the test is designed so that the scores are uniformly distributed between 0 and 100, i.e. $X_i \sim$ Uniform(0,100), so that
- What is the PDF for $Y = \max(X_1, \ldots, X_{10})$?

Note that, since $X_i \sim$ Uniform(0, 100), then for $0 < y < 100$,

$$P(X_i \leq y) = y/100\,.$$

PDF/CDF: Example 2

It is much easier to get the CDF of Y first! Notice that $\max\{X_i\} \leq y$ if and only if $X_i \leq y$ for all X_i. So:

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(X_1 \leq y, X_2 \leq y, \ldots, X_{10} \leq y) \\
&= P(X_1 \leq y) \cdot P(X_2 \leq y) \cdots P(X_{10} \leq y) \\
&= \frac{y}{100} \cdot \frac{y}{100} \cdots \frac{y}{100} \\
&= \frac{y^{10}}{100^{10}}
\end{aligned}$$

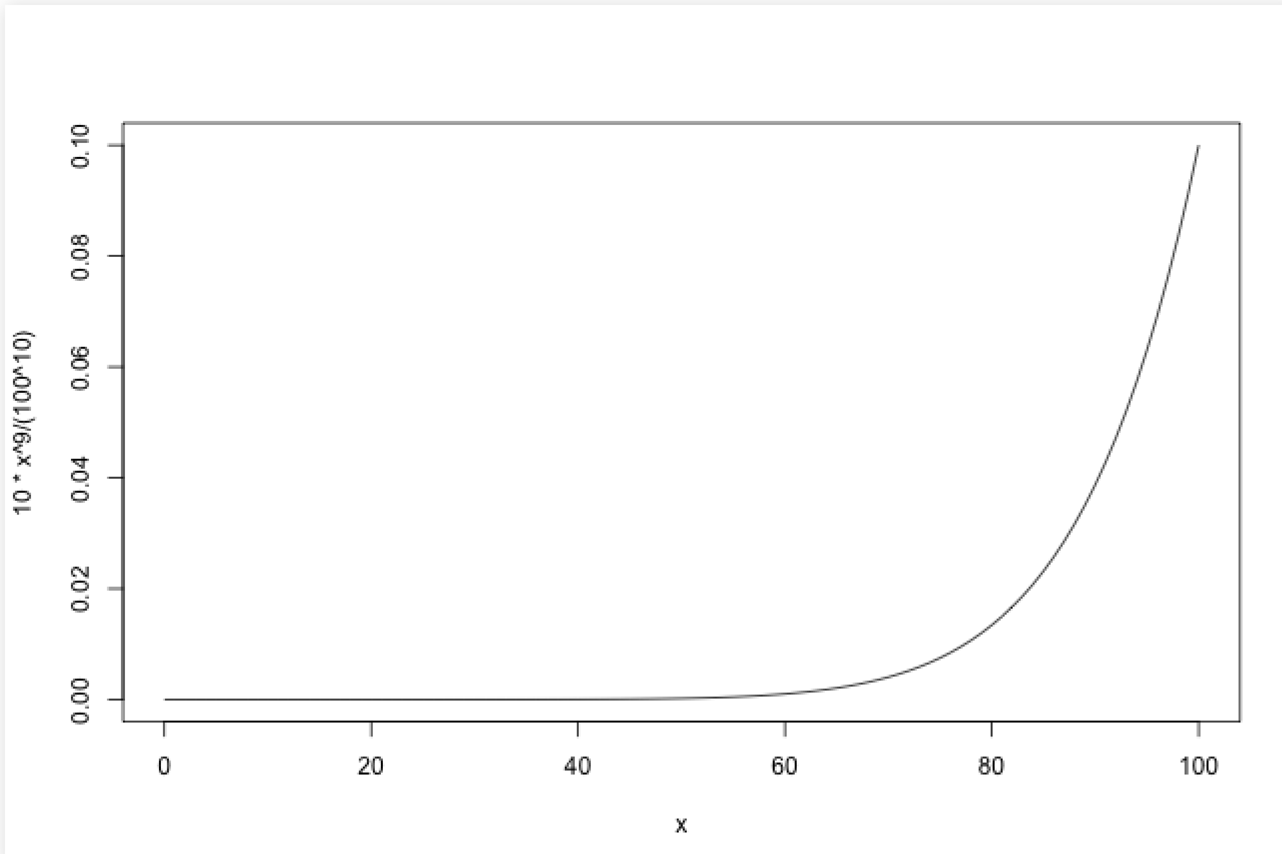
PDF/CDF: Example 2

So the PDF is

$$\begin{aligned} f_Y(y) &= F_Y'(y) \\ &= \frac{d}{dy}\frac{y^{10}}{100^{10}} \\ &= \frac{10y^9}{100^{10}} \end{aligned}$$

PDF/CDF: Example 2

PDF:



Sanity check by Monte Carlo

Let's dive in to `testmax_example.R` on the class website.

Moments for continuous RVs

Remember expected value and variance for discrete random variables. Suppose that X takes the values $x_1, x_2, \ldots, x_N$. Then

$$\mu = E(X) = \sum_{i=1}^{N} x_i \cdot P(X = x_i)$$

and

$$\sigma^2 = \mathrm{var}(X) = \sum_{i=1}^{N} (x_i - \mu)^2 \cdot P(X = x_i)$$

Moments for continuous RVs

In the continuous case, the sum becomes an integral:

$$\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

and

$$\sigma^2 = \mathrm{var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2\, f(x)\, dx$$

A few comments

Warning! Continuous random variables and PDFs can be confusing.

- If X is continuous, then $P(X = x) = 0$ for every point x. *Only sets with nonzero length have positive probability.* If this seems weird, blame it on the real number system.
- The PDF doesn't give you probabilities directly. That is, $f_X(x) \neq P(X = x)$. This only holds for the PMF of a discrete random variable.
- Unlike a PMF, a PDF can be larger than 1. For example, say $f(x) = 3$ for $0 \leq x \leq 1/3$, and $f(x) = 0$ otherwise. This is a well-defined PDF since $\int_{\mathcal{R}} f(x)dx) = 1$.

A few comments

Here a some useful facts about CDFs:

- $P(a < X \leq b) = F_X(b) - F_X(a)$
- $P(X > x) = 1 - F_X(x)$
- If X is continuous, then

$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$$

(Including/excluding endpoints makes no difference.)

The inverse CDF

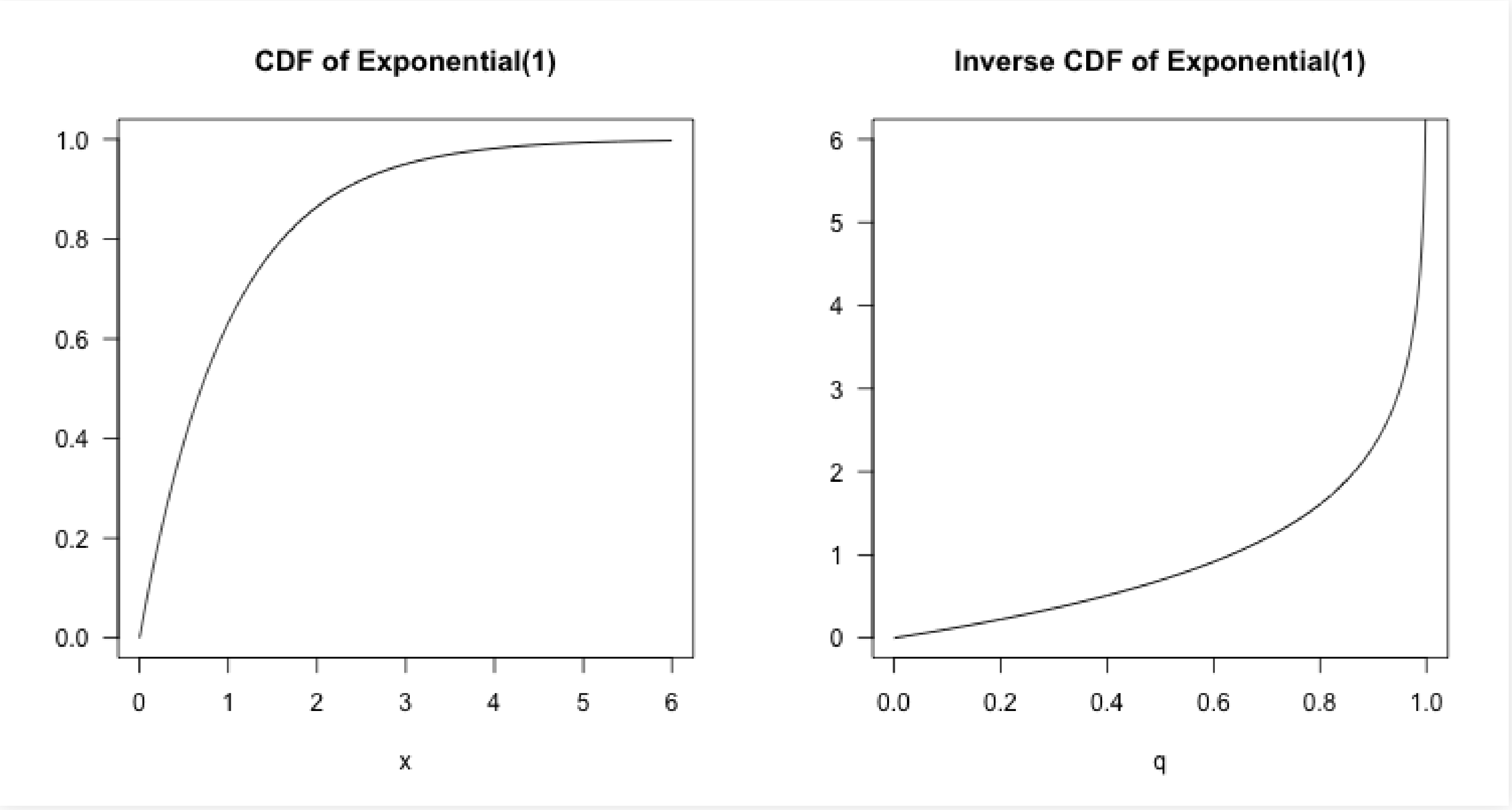
Let X be a random variable with CDF $F_X(x)$, and let $q \in [0, 1]$ be some desired quantile (e.g. 0.9 for the 90th percentile).

If $F(x)$ is continuous and monotonically increasing (i.e. no flat regions), the we define the *inverse CDF* $F^{-1}(q)$, or *quantile function*, as the unique x such that $F(x) = q$.

- $F^{-1}(0.5)$ is the median.
- $F^{-1}(0.25)$ and $F^{-1}(0.75)$ are the first and third quartiles.
- Etc.

The inverse CDF

Example: CDF and inverse CDF of an Exponential(1) random variable.



The inverse CDF

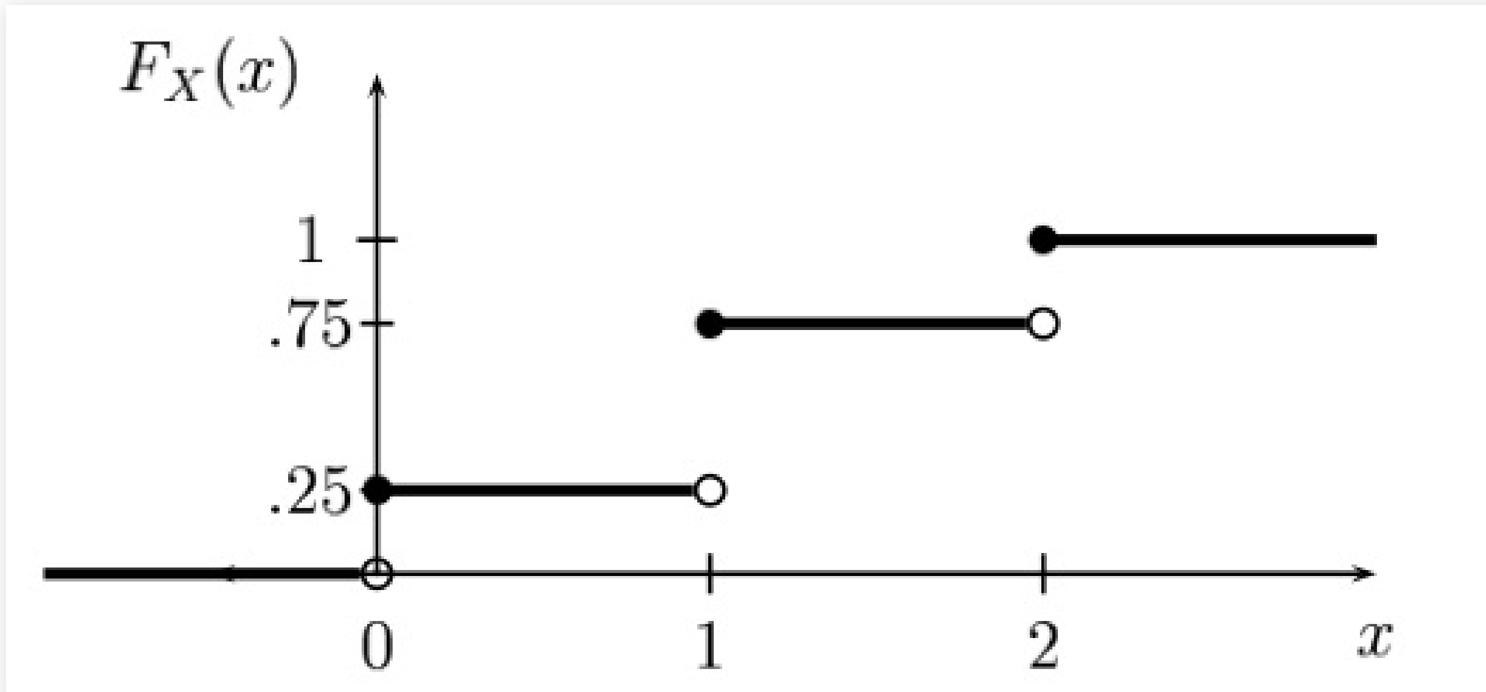
But what if $F(x)$ has flat regions? Then we define

$$F_X^{-1}(q) = \inf \{x : F(x) \geq q\}$$

If you've never seen $\inf$ (infimum) before, just think of it as $\min$. In words, this equation says:

- For a given quantile q, find all the x's for which $F(x) \geq q$.
- Take the smallest such x and call in the inverse.

The inverse CDF



Back to Binomial(N=2, p=0.5). What is $F^{-1}(0.3)$? $F^{-1}(0.75)$? $F^{-1}(0.7501)$?

Normal random variables

A continuous random variable X has a normal distribution if its PDF takes the form

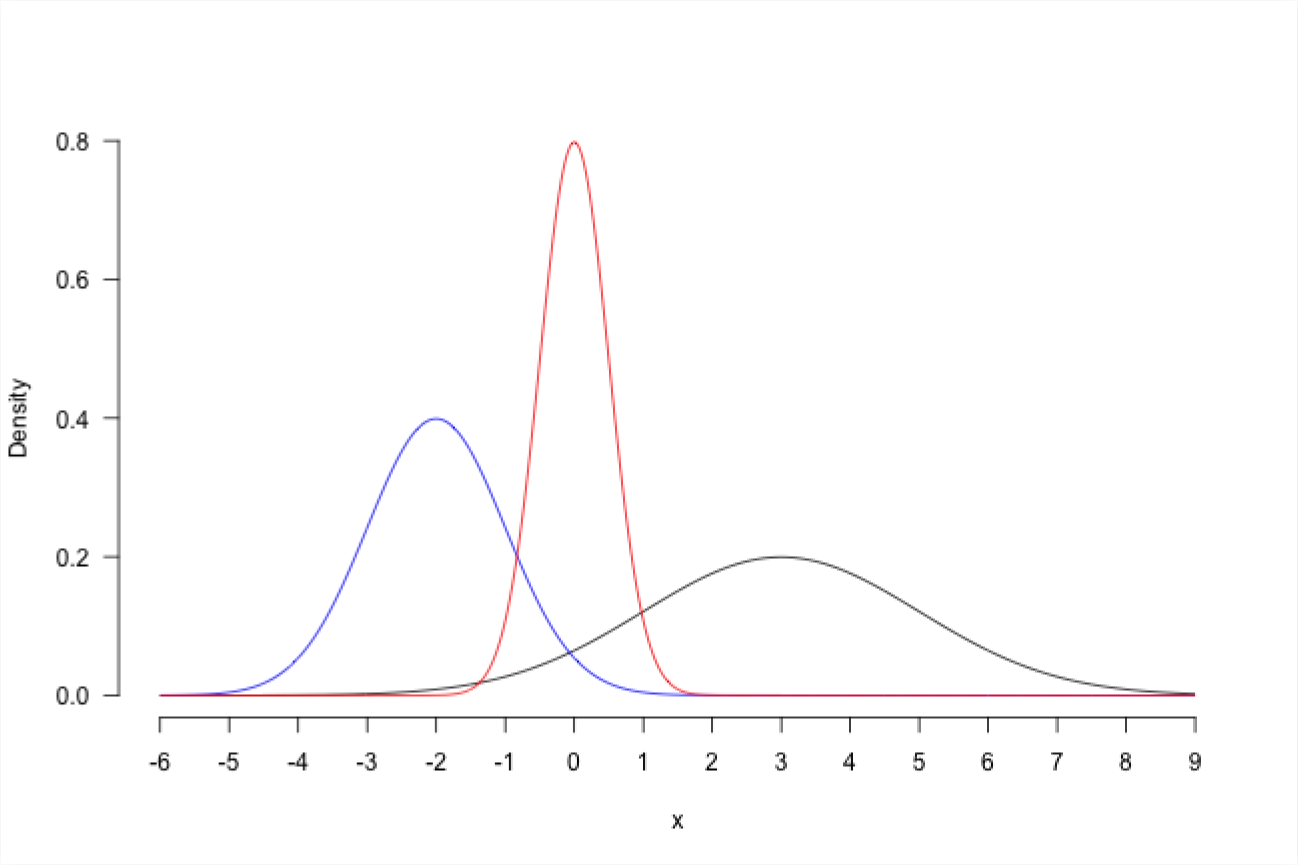
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\}$$

for parameters μ and σ^2.

We write this as $X \sim N(\mu, \sigma)$.

One of the most useful distributions in all of probability and statistics!

Normal random variables



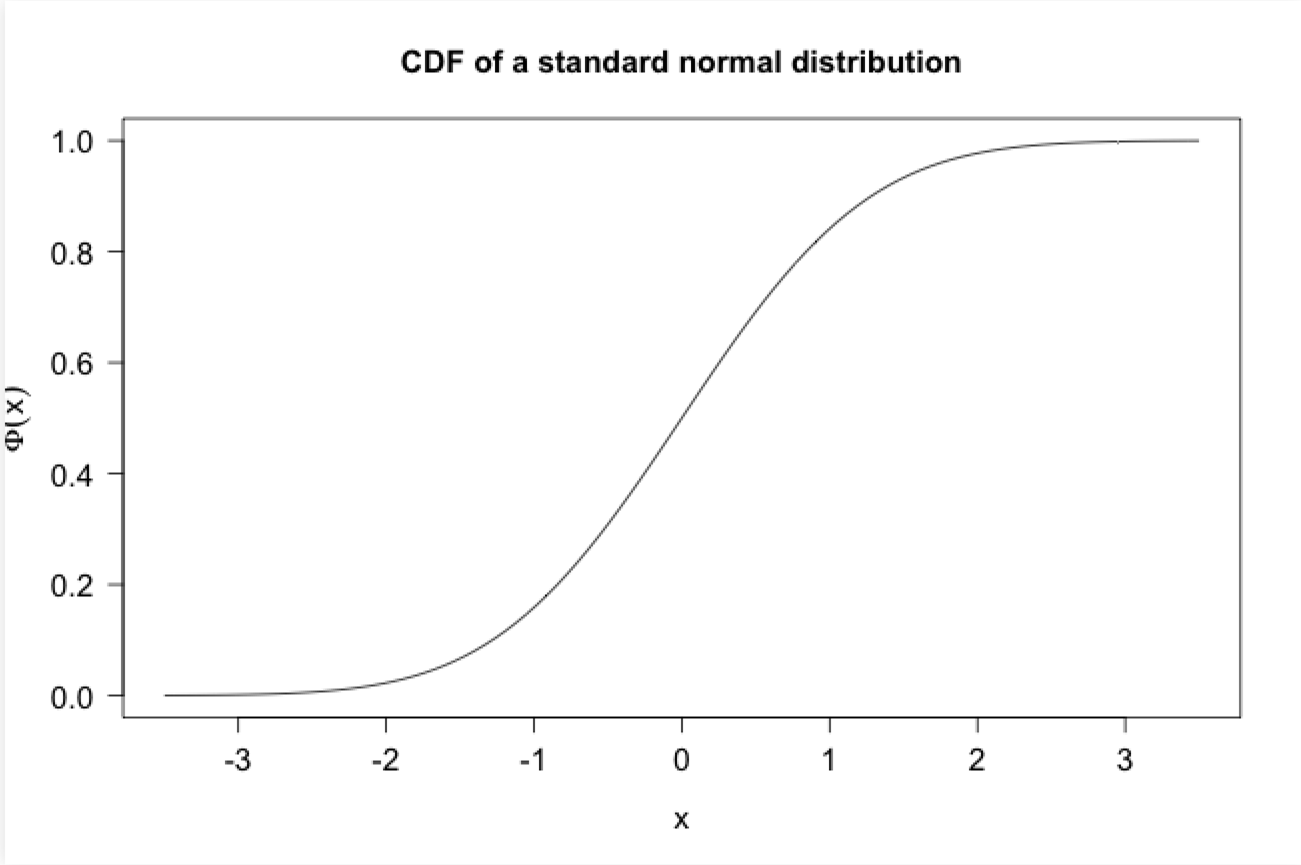
Three different normals: $(\mu = 3, \sigma = 2)$, $(\mu = -2, \sigma = 1)$, and $(\mu = 0, \sigma = 0.5)$

Facts about the normal

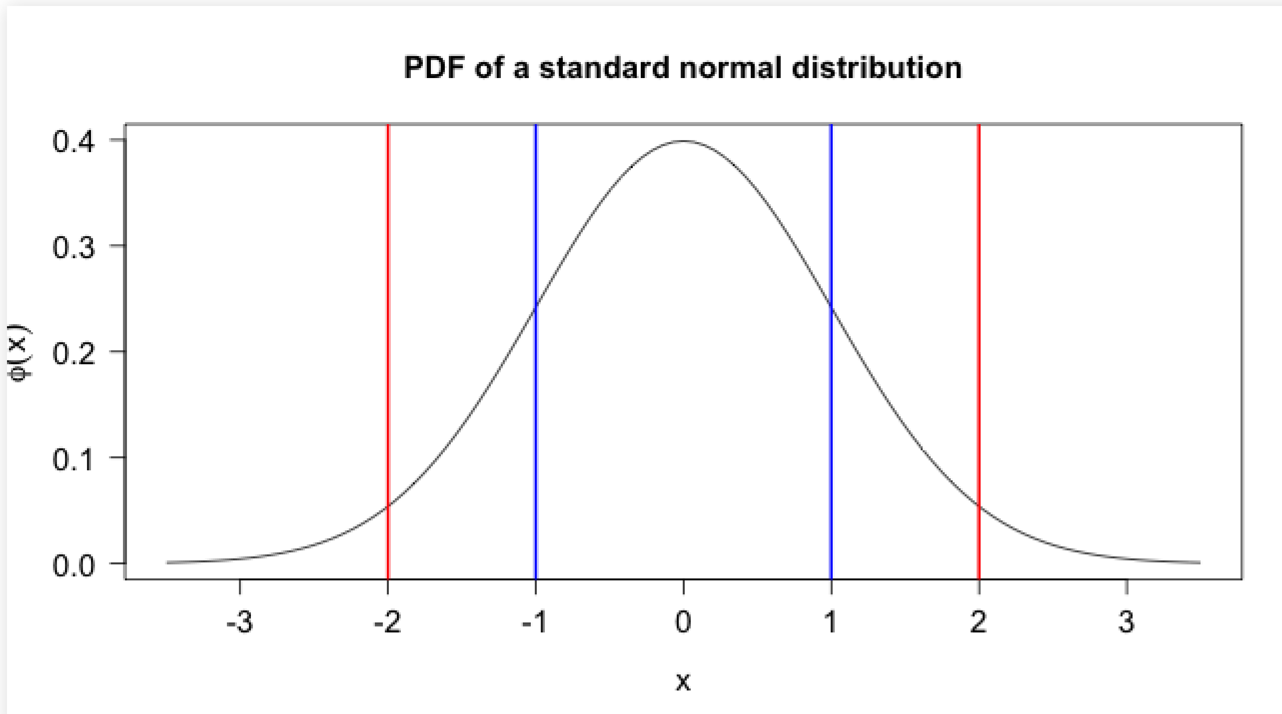
- If $X \sim N(\mu, \sigma)$, then $E(X) = \mu$ and $\mathrm{var}(X) = \sigma^2$.
- If $Z \sim N(0, 1)$, we say that Z has a *standard normal distribution.*
- If $X \sim N(\mu, \sigma)$, and $Z = (X - \mu)/\sigma$, then $Z \sim N(0, 1)$. *We call this "z-scoring."*
- If $Z \sim N(0, 1)$ and $X = \sigma Z + \mu$, then $X \sim N(\mu, \sigma^2)$.
- We use the Greek letter Φ to denote the CDF of a standard normal: if $Z \sim N(0, 1)$, then $\Phi(z) = F_Z(z) = P(Z \leq z)$. There is no closed-form mathematical expression for Φ.

Facts about the normal

Here's a picture of $\Phi(x)$:



Facts about the normal



A couple of useful *critical values*: if Z is standard normal, then

- $P(|Z| \leq 1) \approx 0.68$
- $P(|Z| \leq 2) \approx 0.95$ (More accurately, $P(|Z| \leq 1.96) = 0.9500042$)

Facts about the normal

If $X \sim N(\mu, \sigma)$, then we can obtain the CDF of X from the CDF of a standard normal:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{x-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}$$

where Z is a standard normal random variable.

Facts about the normal

The normal is used *everywhere*:

- distortion of signals in communcations and signal processing
- errors made by a predictive model in statistics and machine learning
- returns on an asset in a financial portfolio
- properties of populations (heights, weights, 100m dash times, etc)

Sometimes sensibly, sometimes inappropriately!

Facts about the normal

The normal distribution is also super important in statistical inference because of something called the **Central Limit Theorem.**

Very roughly, the central limit theorem says that averages of many independent measurements tend to look normally distributed, *no matter what distribution the individual measurements have*.

Facts about the normal

The CLT is one of the deepest and most useful insights in the history of mathematics!

It took over 80 years to work out properly, from de Moivre (1718) to Gauss and Laplace (early 1800s).

It's important in data science because we take averages a lot… and it's therefore important in the real word because so many decisions are made using data!

We'll cover this later.

An example: portfolio returns

Suppose that at age 25, you put $W_0 =$ \$10,000 in the S&P 500, expecting to withdraw it forty years later, when you're 65.

Let W_{40} be the value of your investment after 40 years. What should we expect W_{40} look like?

Classic assumption: suppose that average returns on the stock market, net of inflation, are $r \in (0, 1)$. Here r is like an interest rate: that is, if $r = 0.07$ you average 7% returns a year, and so on.

An example: portfolio returns

Under this assumption we can write W_{40} in terms of formula for compound interest:

$$\begin{aligned} W_{40} &= W_0 \cdot (1 + r) \cdot (1 + r) \cdots (1 + r) \quad \text{(40 times)} \\ &= W_0 \cdot (1 + r)^{40} \end{aligned}$$

So if $r = 0.07$, then

$$W_{40} = 10000 \cdot (1 + 0.07)^{40} = 149744.6$$

An example: portfolio returns

Under this assumption we can write W_{40} in terms of formula for compound interest:

$$
\begin{aligned}
W_{40} &= W_0 \cdot (1+r) \cdot (1+r) \cdots (1+r) \quad \text{(40 times)} \\
&= W_0 \cdot (1+r)^{40}
\end{aligned}
$$

So if $r = 0.07$, then

$$W_{40} = 10000 \cdot (1 + 0.07)^{40} = 149744.6$$

What's wrong with this picture?

An example: portfolio returns

At least two things are wrong:

1. Even if the returns *average* 7%, they won't be 7% every year. Some years will be higher, some lower.
2. We don't know what those returns are going to be! There is tremendous uncertainty.

An example: portfolio returns

Problem 1: the returns fluctuate from year to year.

- So let's explicitly acknowledge that uncertainty in our expression for W_{40}.
- Let R_i be the return on the stock market in year $i = 1, \ldots, 40$.
- Our expression for W_{40} now accumulates these 40 *different* returns over time:

$$
\begin{aligned}
W_{40} &= W_0 \cdot (1 + R_1) \cdot (1 + R_2) \cdots (1 + R_{40}) \\
&= W_0 \cdot \prod_{i=1}^{40} (1 + R_i)
\end{aligned}
$$

An example: portfolio returns

Problem 2: we don't know what the returns will be!

- So let's suppose that $X_i \sim N(\mu, \sigma^2)$.
- We can then use past data to verify whether the normal assumption looks reasonable, and to estimate μ and σ.

Let's dive in to `normal_example.R` on the class website.

New random variables from old ones

Suppose X is a random variable with known probability distribution.

Now we define a new random variable $Y = g(X)$ for some known, fixed g. For example:

- Let X be tomorrow's high temperature in degrees Celsius Then $Y = 1.8X + 32$ is tomorrow's high in degrees Fahrenheit.
- Let X be your average speed in MPH on your 5-mile commute tomorrow. Then $Y = 60 \cdot 5/X$ is the time in minutes it will take you to get to work.

New random variables from old ones

A key question is: what can we say about Y based on what we know about X?

We'll first focus on summaries: $E(Y)$ and $\mathrm{var}(Y)$. In general, expectations and transformations *do not commute*: that is,

- $E(g(X)) \neq g(E(X))$
- $\mathrm{var}(g(X)) \neq g(\mathrm{var}(X))$

Then we'll ask: how can we get a full probability distribution for Y, based on the probability distribution for X?

Expected values under transformation

Linear functions are the one case where the rules for expectation and variance are easy. Suppose that X is some random variable, and that $Y = aX + b$ for constants a and b.

Then

$$
\begin{aligned}
E(Y) &= E(aX + b) = aE(X) + b \quad \text{(Linearity of expectation)} \\
\mathrm{var}(Y) &= a^2 \ \mathrm{var}(X) \\
\mathrm{sd}(Y) &= |a| \ \mathrm{sd}(X)
\end{aligned}
$$

Let's prove this on the board using the definition of expectation.

Expected values under transformation

For nonlinear functions, things are not as nice. To calculate $E(Y)$, we must go back to the PMF/PDF. In words: the expectation of $g(X)$ is the weighted average outcome for $g(X)$, weighted by the probabilities.

If X is discrete with PMF $p_X(x)$:

$$E(Y) = E(g(x)) = \sum_{x \in \mathcal{X}} g(x)\, p_X(x)$$

And if X is continuous with PDF $f_X(x)$:

$$E(Y) = E(g(x)) = \int_{x \in \mathcal{X}} g(x)\, f_X(x)\, dx$$

Expected values under transformation

This rule makes sense, intuitively. Suppose you play a game in Vegas where you draw X randomly from some distribution, and the casino pays you $Y = g(X)$.

Your expected payoff is $g(x)$, times the chance the $X = x$, summed or integrated over all values of X.

New random variables from old ones

OK, what about characterizing the full distribution for $Y = g(X)$?

If X is discrete, things are easy. If $Y = g(X)$, then the PMF of Y can be obtained directly from the PMF of X:

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x)$$

In words: to obtain $p_Y(y)$ we add the probabilities of all values of x such that $g(x) = y$.

New random variables from old ones

The continuous case is harder. There are three steps for finding the PDF $f_Y(y)$ of a transformation $Y = g(X)$.

1. For each y, find the set $A_y = \{x : g(x) \leq y\}$.
2. Find the CDF:

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P(g(X) \leq y) \\ &= P(X \in A_y) \\ &= \int_{x \in A_y} f_X(x)\, dx \end{aligned}$$

3. Compute the PDF as $f_Y(y) = F_Y'(y)$.

Example

What we know: a factory produces oil drills whose diameters average 1 ft, but that have a bit of manufacturing variance.

- Suppose that $X \sim$ Uniform(0.7, 1.3).
- Note: real manufacturing standards are much tighter than this!

What we really care about: oil flows through the resulting borehole at a rate proportional to its cross-sectional area: $Y = \pi \cdot (X/2)^2 = (\pi/4) \cdot X^2$.

What is the probability density of Y? Let's work this together on the board and in `wellbore.R`.