Random Variables

Ben Gravell

benjamin.gravell@utdallas.edu
The Erik Jonsson School of Engineering and Computer Science
The University of Texas at Dallas
800 W. Campbell Rd.
Richardson, TX 75080

Outline

1. Random variables
2. Functions of random variables

Random variables

Random variables

Random variable

A **random variable (RV)** X is a function that maps the sample space Ω to real numbers $\mathbb{R}$ i.e. $X : \Omega \to \mathbb{R}$ that satisfies the following properties:

1. For every Borel set of numbers B, the set $E_B = \{\zeta \in \Omega, X(\zeta) \in B\}$ is an event.
2. $P[X = \infty] = P[X = -\infty] = 0$

Realizations

Upon outcome ζ, a random variable produces a **realization** / **observation** $X(\zeta)$, which is simply a number.

- Think of a realization "popping into being" upon some trigger.
- As shorthand we often refer to the realizations by the same name/variable as the RV.
- We can only observe realizations of the random variable, but not the random variable itself.
- Qualities of the random variable must either be
  1. Assumed before-hand (model)
  2. Inferred from realizations (data)

Flip a coin:
X is one or zero for heads or tails respectively

Roll a die:
X is 1, 2, 3, 4, 5, 6, corresponding to the number of dots on the die face

Spin a wheel:
X is the angle at which it lands between 0 and 360 degrees

Cumulative distribution function (cdf)

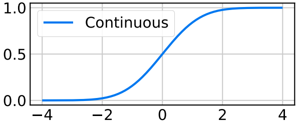
The **cumulative distribution function (cdf)** is defined as

$$F_X(x) = P[\{\zeta | X(\zeta) \leq x\}] \tag{1}$$

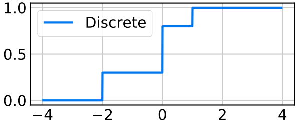
Notation: From here we will usually drop the notation of ζ related to the underlying probability space, so $P[\{\zeta | X(\zeta) \leq x\}]$ becomes $P[X \leq x]$.

Continuous and discrete random variables

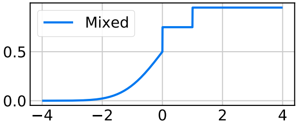
If the cdf $F_X(x)$ is everywhere continuous and differentiable, then X is a **continuous random variable**.



If the cdf $F_X(x)$ is piecewise constant (stairstep shape), then X is a **discrete random variable**.



If neither holds, then X is a **mixed random variable**.



See `mixed.py`

Mass and density functions

Probability mass function (pmf)

The **probability mass function (pmf)** of a discrete random variable is defined as

$$
\begin{aligned}
P_X(x) &= P[X = x] && (2)\\
&= P[X \leq x] - P[X < x] && (3)
\end{aligned}
$$

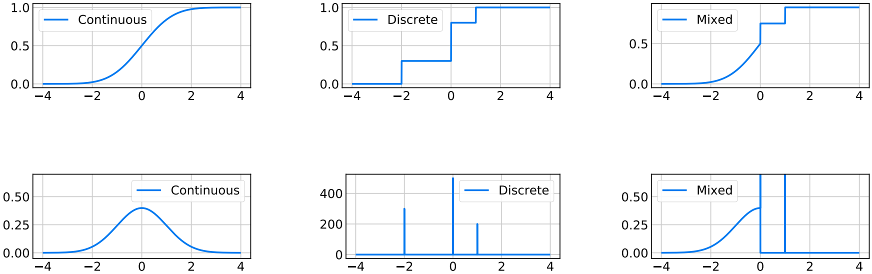
Probability density function (pdf)

The **probability density function (pdf)** of a continuous random variable* is defined as

$$f_X(x) = \frac{d}{dx} F_X(x) \tag{4}$$

* By introducing Dirac delta functions, the pdf can be defined for discrete and mixed random variables.

cdfs on top row, pdfs on bottom row



See `mixed.py`

Properties of the cdf

1. $F_X(\infty) = 1$, $F_X(-\infty) = 0$
2. $F_X(x)$ is nondecreasing in x,
   i.e. $X_1 \leq x_2$ implies $F_X(x_1) \leq F_X(x_2)$
3. $F_X(x)$ is continuous from the right,
   i.e. $F_X(x) = \lim_{\epsilon \to 0^+} F_X(x + \epsilon)$

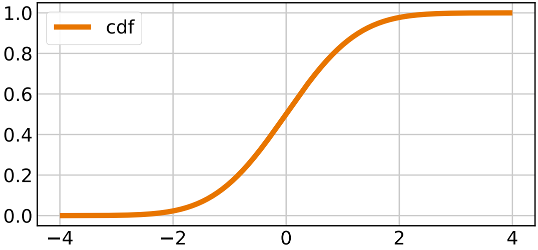


Figure 1: Plot of a typical cdf (std normal)

Properties of the pdf

1. $f_X(x) \geq 0$
2. $\int_{-\infty}^{\infty} f_X(\xi) d\xi = F_X(\infty) - F_X(-\infty) = 1$
3. $F_X(x) = \int_{-\infty}^{x} f_X(\xi) d\xi = P[X \leq x]$
4. $F_X\left(x_2\right) - F_X\left(x_1\right) = \int_{x_1}^{x_2} f_X(\xi) d\xi = P\left[x_1 < X \leq x_2\right]$

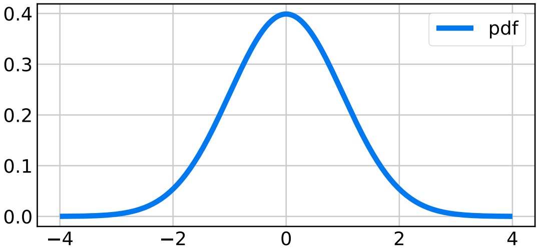


Figure 2: Plot of a typical pdf (std normal)

Knowledge of either the pdf or cdf is sufficient to compute the other, via integration or differentiation.

When we refer to a "distribution," we mean anything that fully specifies a random variable:

- pdf / pmf
- cdf
- Moment generating function (see Ch. 4.5 of [1])
- Characteristic function (see Ch. 4.7 of [1])

Let's introduce a couple of quick concepts before we survey various distributions

Support of a distribution

The **support** of a distribution is the set of values that the random variable X can take with nonzero probability density, i.e.

$$\text{supp}(X) = \{x \mid f_X(x) > 0\}. \tag{5}$$

The distinction between the support and the sample space only comes into effect when the sample space is bigger than required by X

- Sometimes convenient when working with different random variables on a shared sample space
- Example: Two dice with faces $\{1, 1, 1, 2, 3, 3\}$ and $\{2, 3, 4, 5, 6, 6\}$ have different supports $\{1, 2, 3\}$ and $\{2, 3, 4, 5, 6\}$, but we might want a sample space $\{1, 2, 3, 4, 5, 6\}$ to accommodate every possible outcome from either of dice

Mixture distributions

Mixture distribution

A **mixture distribution** is the distribution of a **mixture random variable** Y formed as a composite of other component random variables $X_1, X_2, \ldots, X_N$ by selecting among them at random according to weights $w_1, w_2, \ldots, w_N$.

If the component pdfs are $f_{X_1}, f_{X_2}, \ldots f_{X_N}$, then the **mixture pdf** is simply the weighted average

$$f_Y(Y) = \sum_{i=1}^{N} w_i f_{X_i} \tag{6}$$

Discrete distributions

Trivial distribution

What happens if we treat a non-random, fixed, constant number as a random variable? (w.l.o.g. set $X = 0$)

Trivial distribution

A **trivial** random variable has the pmf

$$P[X = x] = \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

Accordingly, the pdf is the Dirac delta function

$$f_X(x) = \delta(x) \tag{8}$$

and the cdf is the Heaviside step function

$$F_X(x) = H(x) \tag{9}$$

All discrete distributions can be "built" from mixtures of this distribution.

- Follows by definition of pmf

Bernoulli distribution

Bernoulli distribution

A **Bernoulli** random variable has the pmf

$$P[X = x] = \begin{cases} p & \text{if } x = 1, \\ 1 - p & \text{if } x = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

If p is not specified, then assume $p = 1/2$.

Example: A coin flip is Bernoulli where heads $= 1$ and tails $= 0$.

Rademacher distribution

A **Rademacher** random variable has the pmf

$$P[X = x] = \begin{cases} p & \text{if } x = 1, \\ 1 - p & \text{if } x = -1, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Basically just the symmetric version of Bernoulli (which is asymmetric)

- Use whichever is most convenient for the task at hand

Example: A coin flip is Rademacher where heads $= 1$ and tails $= -1$.

Homework P2-1: If X is a Bernoulli random variable, write down a function g such that $Y = g(X)$ is a Rademacher random variable. Also, write down an inverse function $h = g^{-1}$ such that $X = h(Y)$ recovers a Bernoulli distribution. Prove that your functions are correct by directly evaluating the pmfs of $g(X)$ and $h(Y)$.

Consider the binomial experiment with n independent success/fail trials, each governed by a Bernoulli RV.

The number of ways to choose k elements from a population of size n (irrespective of their ordering) is called the number of **combinations** and is determined by the **binomial coefficient**

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \tag{12}$$

The probability of an experiment with k successes and $n-k$ failures is

$$p^k(1-p)^{n-k} \tag{13}$$

Since there are $\binom{n}{k}$ ways in which the experiment could end like this, the probability of seeing an experiment with k successes and $n-k$ failures is

$$\binom{n}{k} p^k(1-p)^{n-k} \tag{14}$$

Binomial distribution

Binomial distribution

A random variable X follows a **binomial distribution** if it represents getting exactly k successes out of the n trials, whose pmf is

$$P[X = k] = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k = 0, 1, \ldots, n, \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

where $p \in [0, 1]$ is a parameter representing the success probability of each trial.

Homework P2-2: (1.56 in [1])
In a particular communication network, the server broadcasts a packet of data to N receivers. The server then waits to receive an acknowledgment message from each of the N receivers before proceeding to broadcast the next packet. If the server does not receive all the acknowledgments within a certain time period, it will rebroadcast (retransmit) the same packet. The server is then said to be in the "retransmission mode." It will continue retransmitting the packet until all N acknowledgments are received. Then it will proceed to broadcast the next packet.

Let $p := P$[successful transmission of a single packet to a single receiver along with successful acknowledgment]. Assume that these events are independent for different receivers and separate transmission attempts. Due to random impairments in the transmission media and the variable condition of the receivers, we have that $p < 1$.

(continued on next slide)

Homework P2-2 (cont.):
(a) In a fixed protocol of method of operation, we require that all N of the acknowledgments be received in response to a given transmission attempt for that packet transmission to be declared successful. Let the event $S(m)$ be defined as follows: $S(m) := \{$ a successful transmission of one packet to all N receivers in m or fewer attempts $\}$.

Find the probability

$$P(m) := P[S(m)]$$

Hint: Consider the complement of the event $S(m)$.

(continued on next slide)

Homework P2-2 (cont.):
(b) An improved system operates according to a dynamic protocol as follows. Here we relax the acknowledgment requirement on retransmission attempts, so as to only require acknowledgments from those receivers that have not yet been heard from on previous attempts to transmit the current packet. Let $S_D(m)$ be the same event as in part (a) but using the dynamic protocol. Find the probability

$$P_D(m) := P[S_D(m)]$$

Hint: First consider the probability of the event $S_D(m)$ for an individual receiver, and then generalize to the N receivers.

(continued on next slide)

Homework P2-2 (cont.):
(c) Compare the performance of the two protocols from parts (a) and (b) by comparing $P(m)$ and $P_D(m)$ for $N = 5$ receivers, $m = 2$ transmission attempts, and success probability $p = 0.9$.

Continuous distributions

Uniform distribution

Uniform distribution

A random variable is **uniform** if the pdf is constant over a finite interval $[a, b]$, i.e. of the form

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

Tail behavior: density drops to zero instantly outside $[a, b]$

- Log density decays "infinitely" fast

Homework P2-3: Derive an expression for the cdf of a uniform random variable.

Gaussian distribution

A random variable X is **Gaussian** or **normal** if it has a pdf of the form

$$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) \tag{17}$$

where μ and σ^2 are parameters (we will define and see later they are the mean and variance).

Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$ is read as "X is distributed according to a normal distribution with mean mu and variance sigma-squared."

Special case: If $\mu = 0$ and $\sigma^2 = 1$, then the distribution is called the **standard normal**.

Tail behavior: log density decays quadratically

See `gaussian.py`

Exponential distribution

A random variable X is **exponential** if it has a pdf of the form

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{18}$$

where $\lambda > 0$ is a parameter.

Homework P2-4: Derive an expression for the cdf of an exponential random variable.

Exponential distributions

Laplace distribution

A random variable X is **Laplace** or **double exponential** if it has a pdf

$$f_X(x) = \frac{1}{2\beta} \exp\left(-\frac{|x-\mu|}{\beta}\right) \tag{19}$$

where μ and β are location and scale parameters.

Notice how similar the Laplace distribution is to a Gaussian

Tail behavior: log density decays linearly - heavier than a Gaussian!

Cauchy distribution

Cauchy distribution

A random variable X is **Cauchy** if it has a pdf of the form

$$f_X(x) = \frac{1}{\pi\gamma}\left(\frac{\gamma^2}{(x - x_0)^2 + \gamma^2}\right) \tag{20}$$

where x_0 and γ are location and scale parameters.

Example: The ratio of two independent normal variables $X = Z_1/Z_2$ is Cauchy

The Cauchy distribution is very bizarre pathological distribution

- It actually has an undefined mean and variance! (discussed later)
- Makes parameter estimation tricky

Tail behavior: log density decays logarithmically - heavier than a Laplace!

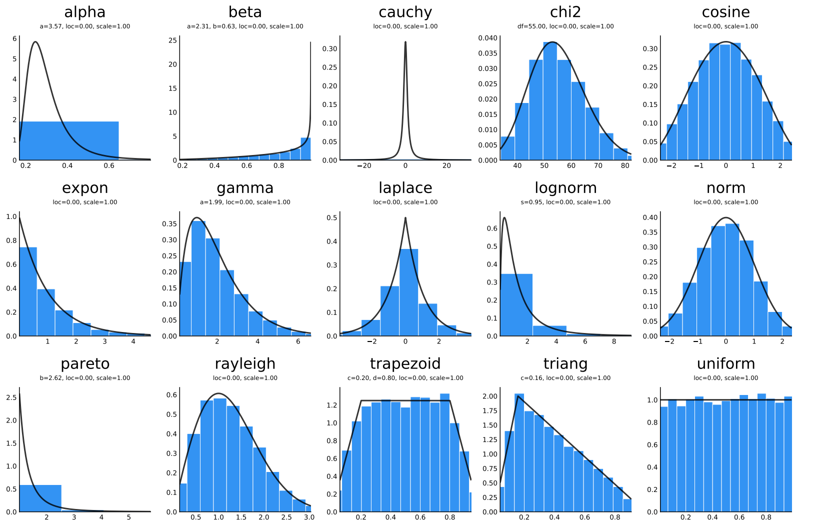


Figure 3: Plot of various pdfs available in SciPy - see `distributions.py`

We can condition random variables on random events

Conditional distribution function

The **conditional distribution function of X given event B** is

$$F_X(x|B) = \frac{P[X \leq x \text{ and } B]}{P[B]} \tag{21}$$

Conditional density function

The **conditional density function of X given event B** is

$$f_X(x|B) = \frac{d}{dx} F_X(x|B) \tag{22}$$

Joint distributions

Just as we had the joint probability of two events, we have the joint distribution of two random variables

Joint distribution function

The **joint (cumulative) distribution function** of X and Y is

$$F_{XY}(x, y) = P[X \leq x \text{ and } Y \leq y] \tag{23}$$

Joint probability mass function

The **joint probability mass function** of X and Y is

$$P_{XY}(x, y) = P[X = x, Y = y] \tag{24}$$

Joint density function

The **joint density function** of X and Y is

$$f_{XY}(x, y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{XY}(x, y) \tag{25}$$

Multinomial distribution

Here is an example of a joint distribution

Idea: Generalize the binomial distribution to trials with more than two outcomes

Consider the multinomial experiment with n independent trials with m outcomes, with each trial governed by a discrete RV with success probabilities $\{p_i\}_{i=1}^m$.

The number of times each outcome happens throughout the entire experiment is a discrete RV X_i for $i = 1, \ldots, m$.

We are interested in the probability that the ith outcome appears exactly k_i times i.e. the joint distribution of the X_i.

Multinomial distribution

The **multinomial coefficient** is the number of ways that the ith outcome appears exactly k_i times (irrespective of their ordering):

$$\frac{n!}{k_1!k_2!\cdots k_m!} \tag{26}$$

The probability of an experiment with the ith outcome appearing exactly k_i times (irrespective of their ordering) is

$$\prod_{i=1}^{m} p_i^{k_i} \tag{27}$$

Since there are $\frac{n!}{k_1!k_2!\cdots k_m!}$ ways in which the experiment could end with the ith outcome appearing exactly k_i times, the probability of seeing such an experiment is

$$\frac{n!}{k_1!k_2!\cdots k_m!}\prod_{i=1}^{m} p_i^{k_i} \tag{28}$$

Multinomial distribution

Multinomial distribution

A collection of RVs $\{X_i\}_{i=1}^m$ follows a **multinomial distribution** if it represents the multinomial experiment, whose joint pmf is

$$
\begin{aligned}
&P[X_1 = k_1, X_2 = k_2, \ldots, X_m = k_m] && (29)\\
&= \begin{cases} \dfrac{n!}{k_1! k_2! \cdots k_m!} \prod_{i=1}^{m} p_i^{k_i} & \text{if } \sum_{i=1}^{m} k_i = n, \\ 0 & \text{otherwise} \end{cases} && (30)
\end{aligned}
$$

where $\{p_i\}_{i=1}^m$ is a set of parameters representing the success probabilities, and must satisfy $\sum_{i=1}^m p_i = 1$.

Exercise: As a special case, how can we recover the binomial distribution from the multinomial distribution?

If we have a joint distribution in hand, we can get the distribution of each of the components by integrating ("marginalizing")

Marginal density function

The **marginal density functions** are

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy \tag{31}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx \tag{32}$$

Marginal distribution function

The **marginal distribution functions** are

$$F_X(x) = F_{XY}(x,\infty) = \int_{-\infty}^{x} f_X(\xi)d\xi \tag{33}$$

$$F_Y(y) = F_{XY}(\infty,y) = \int_{-\infty}^{y} f_Y(\eta)d\eta \tag{34}$$

Conditional distributions

We can also condition random variables on other random variables

Conditional density function

The **conditional density function of X given Y** is

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \tag{35}$$

Conditional distribution function

The **conditional distribution function of X given Y** is

$$F_{X|Y}(x|y) = P[X \leq x | Y \leq y] = \int_{-\infty}^{x} f_{X|Y}(\xi|y) d\xi \tag{36}$$

Notice that $F_{X|Y}(x|y) \neq \frac{F_{XY}(x,y)}{F_Y(y)}$ (unlike the conditional pdf)

Sum Rule for random variables

Let X and Y be two discrete random variables.

The probability that X takes the value x_i, irrespective of the value of Y, is the **total probability** of $X = x_i$, written as $P[X = x_i]$.

Sum Rule for random variables

The total probability of X can be computed as

$$P[X = x_i] = \sum_j P[X = x_i | Y = y_j] P[Y = y_j] \tag{37}$$

$$= \sum_j P[X = x_i, Y = y_j]. \tag{38}$$

This follows from the law of total probability for the event $X = x_i$ and the fact that all the events $Y = y_j$ partition the sample space of Y.

The **total probability** is also referred to as the **marginal probability**, *since we are marginalizing out* the other variable, Y.

Product Rule for random variables

Let X and Y be two discrete random variables.

Conditional probability (Again!)

For only the instances for which $A = a_i$, the fraction of such instances for which $B = b_j$ is $P[B = b_j | A = a_i]$ and are called the **conditional probability of $B = b_j$ given $A = a_i$**.

Product Rule for random variables

The joint pmf of X and Y can be computed as

$$P[X = x_i, Y = y_j] = P[Y = y_j | X = x_i] P[X = x_i] \tag{39}$$

Bayes' theorem for random variables

RV conditioned on RV

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} \tag{40}$$

Event conditioned on RV

$$P[A|X = x] = \frac{f_{X|A}(x)P[A]}{f_X(x)} \tag{41}$$

RV conditioned on event

$$f_{Y|A}(y) = \frac{P[A|Y = y]f_Y(y)}{P[A]} \tag{42}$$

Independent random variables

Independent random variables

Two random variables X and Y are **statistically independent** if the two events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent for any pair (x, y).

Equivalently,

$$F_{XY}(x, y) = F_X(x) F_Y(y) \tag{43}$$

or

$$f_{XY}(x, y) = f_X(x) f_Y(y) \tag{44}$$

You can imagine the generalization to more than two RV's - joint distribution is equal to product of the marginals

It is nice when RV's are independent because it makes computing their joint distribution trivial - just multiply the marginals!

Functions of random variables

Core problem:
What is the distribution of a function of a random variable?

Math:
Given $f_X(x)$ and $Y = g(X)$, what is $f_Y(y)$?

"Indirect" procedure:

1. Find the point set C_y such that $\{Y \le y\} = \{X \in C_y\}$
2. Find the cdf of Y as

$$F_Y(y) = P[Y \le y] = P[g(X) \le y] = P[X \in C_y] \tag{45}$$

3. Find the pdf of Y as

$$f_Y(y) = \frac{d}{dy} F_Y(y) \tag{46}$$

Suppose g is affine, i.e. $Y = g(X) = aX + b$.

<u>Case 1: $a > 0$</u>

Step 1: Find the point set

$$\{Y \leq y\} = \{aX + b \leq y\} \tag{47}$$

$$= \left\{X \leq \frac{y-b}{a}\right\} = \{X \in C_y\} \tag{48}$$

Step 2: Find the cdf

$$F_Y(y) = P[Y \leq y] = P[aX + b \leq y] \tag{49}$$

$$= P\left[X \leq \frac{y-b}{a}\right] = F_X\left(\frac{y-b}{a}\right) \tag{50}$$

Example: affine function of a random variable

Step 3: Differentiate cdf to get pdf
Use the change of variables $z = \frac{y-b}{a}$ so

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) \tag{51}$$

$$= \frac{dF_X(z)}{dz} \cdot \frac{dz}{dy} \qquad \text{(chain rule)}$$

$$= f_X(z) \cdot \frac{1}{a} \tag{52}$$

Optional Exercise: Work out Case 2: $a < 0$

After doing that, you will find the solution is

$$\boxed{f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \quad \text{if } a \neq 0} \tag{53}$$

Optional Exercise: Work out Case 3: $a = 0$ (degenerate case)
Hint: The solution is trivial: $f_Y(y) = \delta(y - b)$*, a Dirac delta at* b*.*

Example 3.2-8 in [1]

Consider the vertical coordinate of a spinner with uniform random angle

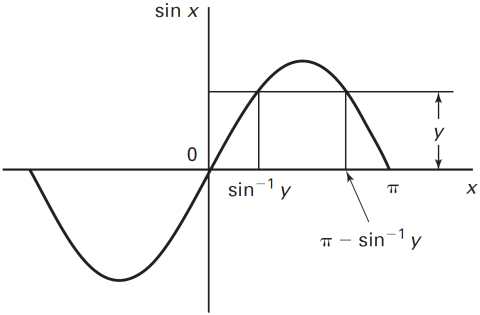
$$g(X) = \sin(X) \quad \text{(sine map)} \tag{54}$$

$$f_X(x) = \begin{cases} \frac{1}{2\pi} & \text{if } -\pi \leq X \leq \pi \\ 0 & \text{else} \end{cases} \qquad \text{(uniform distribution)} \tag{55}$$

Case 1: $0 \leq y < 1$

Step 1: Find the point set (this time it's trickier)

$$\{Y \leq y\} = \{\sin(X) \leq y\} \tag{56}$$
$$= \left\{-\pi < X \leq \sin^{-1}(y)\right\} \cup \left\{\pi - \sin^{-1}(y) < X \leq \pi\right\} \tag{57}$$
$$= \{X \in C_y\} \tag{58}$$



Step 2: Find the cdf

$$F_Y(y) = P[Y \le y] \tag{59}$$
$$= P\left[\left\{-\pi < X \le \sin^{-1}(y)\right\} \cup \left\{\pi - \sin^{-1}(y) < X \le \pi\right\}\right] \tag{60}$$
$$= P\left[-\pi < X \le \sin^{-1}(y)\right] + P[\pi - \sin^{-1}(y) < X \le \pi] \tag{61}$$
$$= \left[F_X(\sin^{-1}(y)) - F_X(-\pi)\right] + \left[F_X(\pi) - F_X(\pi - \sin^{-1}(y))\right] \tag{62}$$

Step 3: Differentiate cdf to get pdf

$$f_Y(y) = \frac{d}{dy} F_Y(y) \tag{63}$$
$$= f_X\left(\pi - \sin^{-1} y\right) \frac{1}{\sqrt{1-y^2}} + f_X\left(\sin^{-1} y\right) \frac{1}{\sqrt{1-y^2}} \tag{64}$$
$$= \frac{1}{\pi} \cdot \frac{1}{\sqrt{1-y^2}} \quad \text{for } 0 \le y < 1 \tag{65}$$

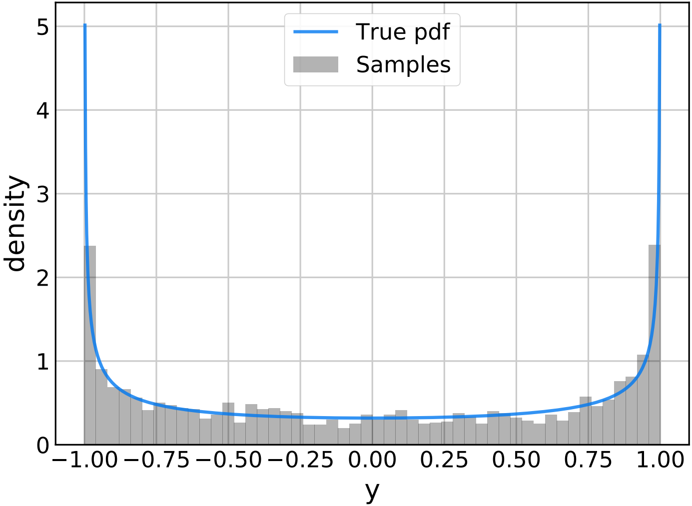
Optional Exercise: Work out Case 2: $-1 < y \leq 0$
Hint: You should find the pdf is the same as for $0 \leq y < 1$

Optional Exercise: Work out Case 3: $|y| >= 1$
Hint: You should find the cdf is constant with respect to y
(either $F_Y(y) = 0$ *or* $F_Y(y) = 1$*) and therefore the pdf is zero.*

Therefore, the complete solution is

$$\boxed{f_Y(y) = \begin{cases} \dfrac{1}{\pi} \cdot \dfrac{1}{\sqrt{1-y^2}} & \text{if } |y| < 1 \\ 0 & \text{else} \end{cases}} \tag{66}$$

We can check our solution against a histogram of empirical samples
- see `function_of_rv.py`



Can we go directly from pdf of X to pdf of $Y = g(X)$ (without finding intermediate cdf)?

“Direct” procedure:

1. Find the root functions $x_i = x_i(y)$ that satisfy $y - g(x_i) = 0$ for any fixed y
2. Compute derivative $g'(x)$
3. Evaluate $|g'(x_i)|$ check $|g'(x_i)| \neq 0$
4. Compute the pdf directly as

$$f_Y(y) = \sum_i \frac{f_X(x_i)}{|g'(x_i)|} \tag{67}$$

Note: Throughout keep in mind that $x_i = x_i(y)$ are functions!

Example 3.2-9 in [1]

Consider again the problem

$$g(X) = \sin(X) \quad \text{(sine map)} \tag{68}$$

$$f_X(x) = \begin{cases} \dfrac{1}{2\pi} & \text{if } -\pi \leq X \leq \pi \\ 0 & \text{else} \end{cases} \quad \text{(uniform distribution)} \tag{69}$$

Case 1: $0 \leq y < 1$

Step 1:
For any $0 \leq y < 1$ we have the roots of

$$y - g(x) = y - \sin(x) = 0 \tag{70}$$

are

$$x_1 = \sin^{-1}(y) \quad \text{and} \quad x_2 = \pi - \sin^{-1}(y) \tag{71}$$

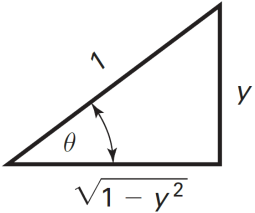
Step 2:
We have the derivative

$$\frac{dg}{dx} = \cos(x) \tag{72}$$

Step 3:
Evaluated at the roots, the derivative is

$$\left.\frac{dg}{dx}\right|_{x_1} = \cos(\sin^{-1}(y)), \qquad \left.\frac{dg}{dx}\right|_{x_2} = -\cos(\sin^{-1}(y)) \tag{73}$$

When you see the **composition of trig and inverse trig**, there is usually a nice simplification to make - use triangle diagram to help



$$\sin(\theta) = \frac{y}{1} \tag{74}$$

$$\theta = \sin^{-1}(y) \tag{75}$$

$$\cos(\theta) = \frac{\sqrt{1-y^2}}{1} \tag{76}$$

$$\cos(\sin^{-1}(y)) = \sqrt{1-y^2} \tag{77}$$

We have the absolute values

$$\left|\frac{dg}{dx}\right|_{x_1} = \left|\frac{dg}{dx}\right|_{x_2} = \sqrt{1-y^2} \neq 0 \text{ for } 0 \leq y < 1 \tag{78}$$

Step 4:
Compute the pdf

$$f_Y(y) = \sum_i \frac{f_X(x_i)}{|g'(x_i)|} \tag{79}$$

$$= \frac{\frac{1}{2\pi}}{\sqrt{1-y^2}} + \frac{\frac{1}{2\pi}}{\sqrt{1-y^2}} \tag{80}$$

$$= \frac{1}{\pi}\sqrt{1-y^2} \quad \text{for } 0 \leq y < 1 \tag{81}$$

which is the same result as we got using the "indirect" method.

Optional Exercise: Repeat the procedure for Case 2: $-1 < y \leq 0$

Optional Exercise: Repeat the procedure for Case 3: $|y| \geq 1$

Core problem:
What is the distribution of a function of a random variable?

Math:
Given $f_{XY}(x, y)$ and $Z = g(X, Y)$, what is $f_Z(z)$?

"Indirect" procedure:

1. Find the point set C_z such that $\{Z \leq z\} = \{(X, Y) \in C_z\}$
2. Find the cdf of Z as

$$F_Z(z) = \iint_{(x,y)\in C_z} f_{XY}(x, y)dxdy \tag{82}$$

3. Find the pdf of Z as

$$f_Z(z) = \frac{d}{dz} F_Z(z) \tag{83}$$

Optional Exercise: Find $f_Z(z)$ where $Z = XY$
Hint: See Example 3.3-1 in [1]
Solution:

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|y|} f_{XY}(z/y, y) dy \tag{84}$$

Optional Exercise: Find $f_Z(z)$ where $Z = X + Y$ Eqs. (3.3-13), (3.3-14) in [1]

Solution:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(z-y, y)dy \tag{85}$$

If X and Y are independent

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)dy \tag{86}$$

which is a **convolution integral**

Evaluate by reversing one function and sliding it

See Examples 3.3-4, 3.3-5, 3.3-6, 3.3-7, 3.3-8 in [1]

Homework: Max of two RVs

Homework P2-4: Find $f_Z(z)$ where $Z = \max(X, Y)$ and X, Y are independent.
Hint: See Example 3.3-2 in [1]

[1] John Woods and Henry Stark.
Probability, Statistics, and Random Processes for Engineers.
Pearson Higher Ed, 4 edition, 2011.