# Mini-HW 15
# 97.3% of all statistics are made up

15-462/662

## 1 Continuous Probability

You're probably (pun intended) familiar with the concept of discrete probability—each possible outcome has some non-zero chance of happening, the chance of multiple disjoint events occurring is the sum of their probabilities, etc. However, unless you've taken a probability/stats class, you might not be familiar with continuous probability.

A continuous random variable $X$ represents the result of some random process that can produce one of an uncountably infinite number of possible outcomes. For example, we write $X \sim \text{Uniform}(0, 1)$ to denote that $X$ has equal chance of being drawn from anywhere in the real interval $[0, 1]$.

We can represent the distribution of possible outcomes for a continuous random variable using a probability density function (PDF) $f : \Omega \mapsto \mathbb{R}$, where $\Omega$ is the set of outcomes. For example, $f_X(x) = 1$ is the probability density function for $X \sim \text{Uniform}(0, 1)$.

A probability density function is similar to the discrete probability *mass* function (PMF), which assigns a non-negative probability to each possible outcome. However, there is one key difference: when the outcomes are drawn from a continuous space, it no longer makes sense for $f_X(x)$ to be the probability of drawing $x$. Since there are uncountably infinite possible values of $x$, if we try to add up all their probabilities, we'll get infinity. Instead, $f_X(x)$ represents the probability *density* at $x$, and the corresponding continuous operation is to *integrate*.

To compute the probability that $X$ falls within some interval $[a, b]$:

$$\mathbf{P}\{X \in [a, b]\} = \int_a^b f_X(x)dx$$

As a distribution, $f_X(x)$ should integrate to exactly 1 over its domain.

### 1.1 Questions

Consider the random variable $X \in [0, 1]$ with distribution $f_X(x) = 2 - 2x$.

1. What is $\mathbf{P}\{X = 0.5\}$?
2. What is $\mathbf{P}\{0.49 \leq X \leq 0.51\}$?
3. A discrete probability distribution is described by a probability *mass* function rather than probability *density* function. Why do these terms differ?

## 2 Sampling

In many applications, we may want to draw samples from a continuous distribution according to some PDF. However, it's not immediately clear how to do so—computers, if they can generate truly random samples at all, can only really sample uniformly. Thankfully, there are two particularly useful methods for generating any distribution given only a uniform sample generator.

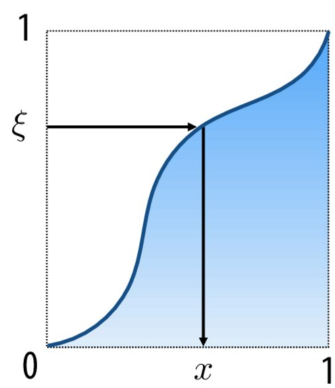
### 2.1 Inversion Sampling

If you have an analytical description of the PDF (or PMF) you wish to sample from, you can use the method of inversion sampling. This method relies on the concept of the cumulative distribution function (CDF), defined as:

$$F_X(x) = \mathbf{P}\left\{X \leq x\right\} = \int_{-\infty}^{x} f_X(z)dz$$

That is, $F_X(x)$ represents the probability that the outcome is at most $x$. If $X$ is defined over the interval $[0, 1]$, then clearly we would have that $F_X(0) = 0$ and $F_X(1) = 1$, as there is no chance that $X = 0$ and 100% chance that $X \leq 1$. Further, since $f_X(x) \geq 0$, we know $F_X(x)$ is non-decreasing.

Given these properties, we may consider the inverse CDF, $F_X^{-1}(x)$. If we plug in a uniformly distributed $0 - 1$ variable to the inverse, we will be uniformly sampling the range of the CDF and getting back the corresponding point in the CDF domain. This is, in fact, sampling the domain with a distribution proportional to the derivative of $F_X(x)$, which we can see pictorially:




...as the amount of the range taken up by a small neighborhood around $x$ is proportional to the slope of the CDF at $x$. Conveniently, it turns out $\frac{d}{dx}F_X(x) = f_X(x)$, so this is sampling exactly the distribution we wanted.

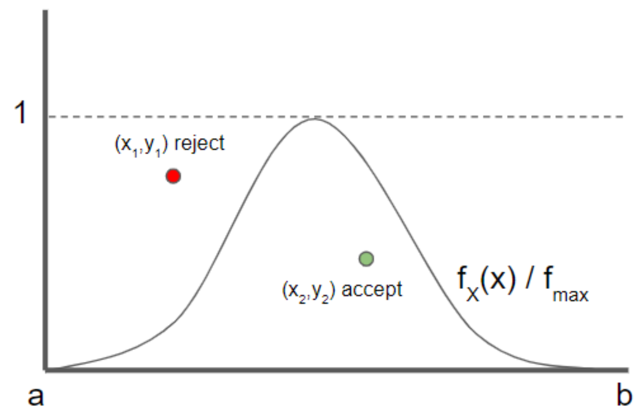
Note that while here we assumed $F_X(x)$ has a unique inverse, this assumption is not actually required. It may be illustrative to think about why.

## 2.2 Rejection Sampling

On the other hand, it might be the case that it's difficult or impossible to analytically find, invert, and evaluate the CDF of your distribution. In this case, you might want to use rejection sampling, as long as your desired distribution is bounded.

Implementing rejection sampling is usually quite simple. First, we need to know the maximum value attained by $f_X(x)$ (precisely $\sup\{f_X(x) \mid x \in X\}$) so that we can later normalize the PDF. Denote the maximum as $f_{max}$. Second, we can uniformly sample $x$ from the domain and accept it with probability $\frac{f_X(x)}{f_{max}}$. If we instead reject the sample, we start over with a new uniform sample of the domain.

In the 1D case, we can also think of this geometrically:




...where we sample some point $(x, y) \in [a, b] \times [0, 1]$, and reject it if $\frac{f_X(x)}{f_{max}} < y$.

To see that this works, consider the probability of ending up with any outcome $x$ (or in the continuous case, a small neighborhood around $x$). If we choose some $x$ in step 1, we will then accept it with probability proportional to $f_X(x)$. If we reject the sample, we have an equal chance of choosing each $x$ for the next attempt, so rejections do not effect the distribution. You can also convince yourself that dividing by $f_{max}$ will properly normalize the resulting PDF. (For proof, refer to this link.)

## 2.3 Questions

For the following procedures, assume you have access to RNG::Unit(), a function which returns a float uniformly distributed in $[0, 1]$.

Consider the random variable $X \in [0, 1]$ with distribution $f_X(x) = 2 - 2x$.

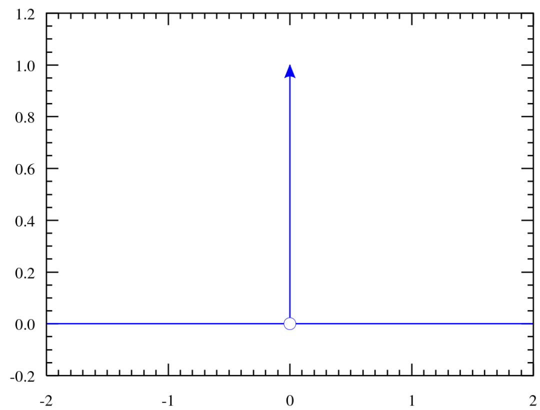
1. Write a procedure that inversion samples $X$.
2. Write a procedure that rejection samples $X$.
3. Which of these methods might require more computation, and which might require more random numbers?
4. (Extra Credit) Prove that $F_X(X) \sim \text{Uniform}(0, 1)$ for any $X$ by showing they have the same CDF. Argue that you can choose some $\widetilde{F_X}^{-1}(X)$ if a unique inverse does not exist.

## 3 Extra Note: Delta Distributions

A special distribution that comes up when discussing BSDFs is the Dirac delta distribution. This distribution $\delta_c(x)$ is non-zero at exactly one point in the domain, but still integrates to 1. Technically, this means:

$$\delta_c(x) = \begin{cases} \infty & x = c \\ 0 & \text{otherwise} \end{cases}$$

such that $\int_\Omega \delta_c(x)dx = 1$. Here's a graph of $\delta_0(x)$:




However, this description is not very illustrative. We can instead think of $\delta$ as an 'if' statement: if $x = c$, then accept, otherwise reject. If we consider a $\delta_c$ as a probability distribution, this implies $\mathbf{P}\{X = c\} = 1$, as the only accepted outcome is $c$. So then, how do we sample $\delta_c$? Simply return $c$ every time.

Hence, we can use the $\delta$ distribution to emulate a discrete PMF using a continuous PDF. For example, consider $f_X(x) = \frac{1}{2}\delta_a(x) + \frac{1}{2}\delta_b(x)$. This continuous PDF emulates the discrete distribution $\mathbf{P}\{X = a\} = \mathbf{P}\{X = b\} = \frac{1}{2}$. If we wanted to sample this, we would simply return $a$ or $b$ with equal probability.

Unfortunately, in this case we still have that $f_X(x)$ is still either 0 or $\infty$, so it is rather hard to compute with. However, if we pretend that our PDF is actually a PMF, it may make sense to return the straight-up probability, which is how Scotty3D implements delta-based BSDFs. The idea of emulating a PMF using deltas will be particularly useful for the glass BSDF.