Moments of a Distribution (4 Questions)

Question 1

The example of the probability integral transformation given in class demonstrates which of the following for continuous random variables?

- ◯ If X is a uniformly-distributed random variable, then the CDF is also uniformly distributed
- ◉ The result that if you transform a random variable by its own CDF, the resulting distribution will be uniform [0,1]
- ◯ That the PDF and the CDF are equivalent functions for uniformly-distributed random variables
- ◯ The nature of the relationship between the PDF and the CDF for all types of distributions

✔

Explanation
In the lecture we learned that transforming a continuous random variable by its CDF yields a random variable that is uniformly distributed. The answer choices about the uniformly distibuted functions are wrong, most easily seen from the fact that the uniform PDF is flat whereas the uniform CDF is a step function.

Question 2

This question is about using the probability integral transform for random number sampling on a computer (for deep reasons outside the scope of this course, random sampling on a computer is necessarily pseudo-random rather than perfectly random, but we will gloss over this minor point).

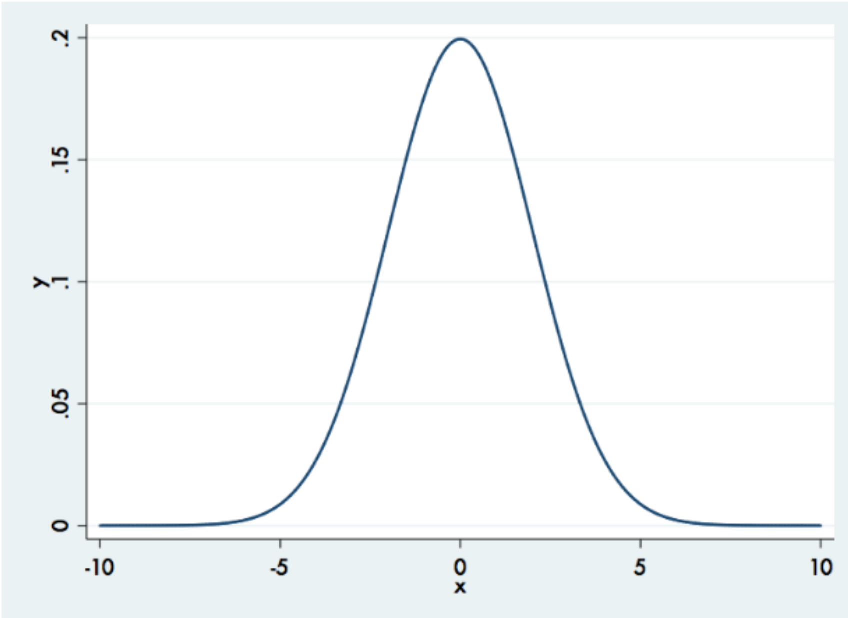
In the general setup, we would like to sample random numbers per a continuous random variable X with known cdf $F(x)$ (or known pdf $f(x)$), given only that you can sample random numbers from the standard uniform $U[0,1]$ distribution (as Prof. Ellison has mentioned, it is typically the case that any statistical package allows you to do this at minimum).

Let's walk through an example problem. Suppose you want to sample a single random number r from the *standard normal distribution* , a distribution which we will examine in more detail later in this course. This distribution has the following pdf:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-0.5x^2}$$

For our purposes, though, all we need are the basic facts of the the graph of the pdf: it is a bell shaped function, symmetric about $x = 0$, with infinite support (it has a nonzero probability density everywhere, but with vanishingly small probability density in the "tails").

Moments of a Distribution (4 Questions)



So to sample a random number from this standard normal distribution, you begin by sampling from the $U[0,1]$ distribution; suppose the number $u \in [0,1]$ you retrieve from the uniform random generator it is exactly 0.500. What should you therefore report as your number r sampled from the standard normal distribution? (Hint: for this particular example, no complex calculations are required!)

- ◯ 1
- ◉ 0
- ◯ 0.5

✔

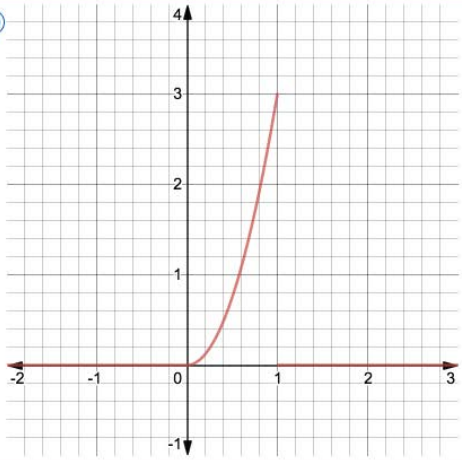
Explanation

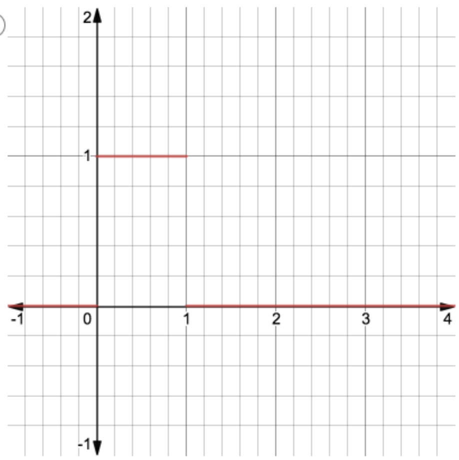
By the probability integral transform method, we want to sample r as $r = F^{-1}(u) = F^{-1}(0.500)$, where F is the cdf of the standard normal distribution. F could be a complicated function, but since the standard normal distribution pdf is symmetric about x = 0, we know $F^{-1}(0.500) = F^{-1}\left(\frac{1}{2}\right) = 0$

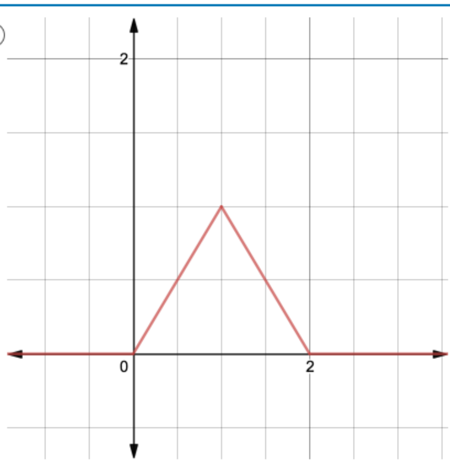
Moments of a Distribution (4 Questions)

Question 1

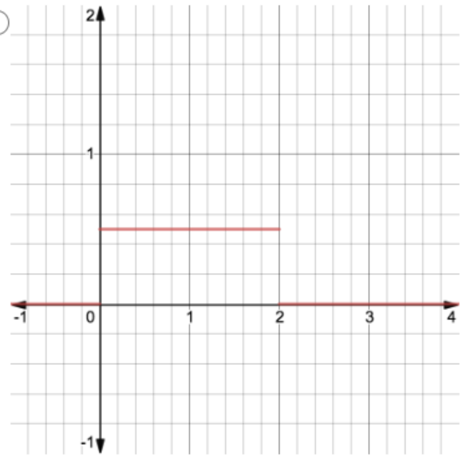
Consider the pdfs for the continuous random variables below. For which of these is the mean possibly not equal to the median?







Moments of a Distribution (4 Questions)



✔

Explanation
For any symmetric distribution (with respect to the nonzero support of the distribution), the mean is always equal to the median - intuitively the mean captures the 'center of mass' of the distribution, and for a symmetric distribution this is exactly the point that divides the pdf into two equal areas (which characterizes the median). Thus it is only in asymmetric distributions that the mean is possibly not equal to the median (though not guaranteed to be inequal - whether or not the mean differs from the median ultimately depends on the exact shape of the distribution).

Question 2

Say you want to find the probabilities $P(a < X < b)$ for any $a < b$ and suppose you only have one of the following pieces of information. Which of these will provide you enough information to find the probabilities? (Select all that apply)

- [ ] CDF ✔
- [x] PDF ✔
- [ ] Mean
- [ ] Median

Explanation
One can always find this from the CDF: If we knew the CDF function $F(x)$, we can always calculate $P(a < X < b) = F(b) - F(a)$.
The PDF tells you either the point probabilities (for a discrete variable) or their densities (for a continuous variable), so they can always be summed or integrated to find the corresponding CDF which can then be used to calculate $P(a < X < b)$.
Neither the mean nor the median provide sufficient information. One can imagine two PDFs $f(x)$ and $g(x)$ with the same mean and median (e.g. two symmetric bell curves centered at 0), except that $f(x)$ is concentrated at its mean/median (e.g. a bell curve with a small standard deviation) and $g(x)$ is more spread out (e.g. a bell curve with a large standard deviation). These PDFs will calculate different cdfs $F(x)$ and $G(x)$, so For fixed a, b, will calculate different probabilities $P(a < X < b)$.