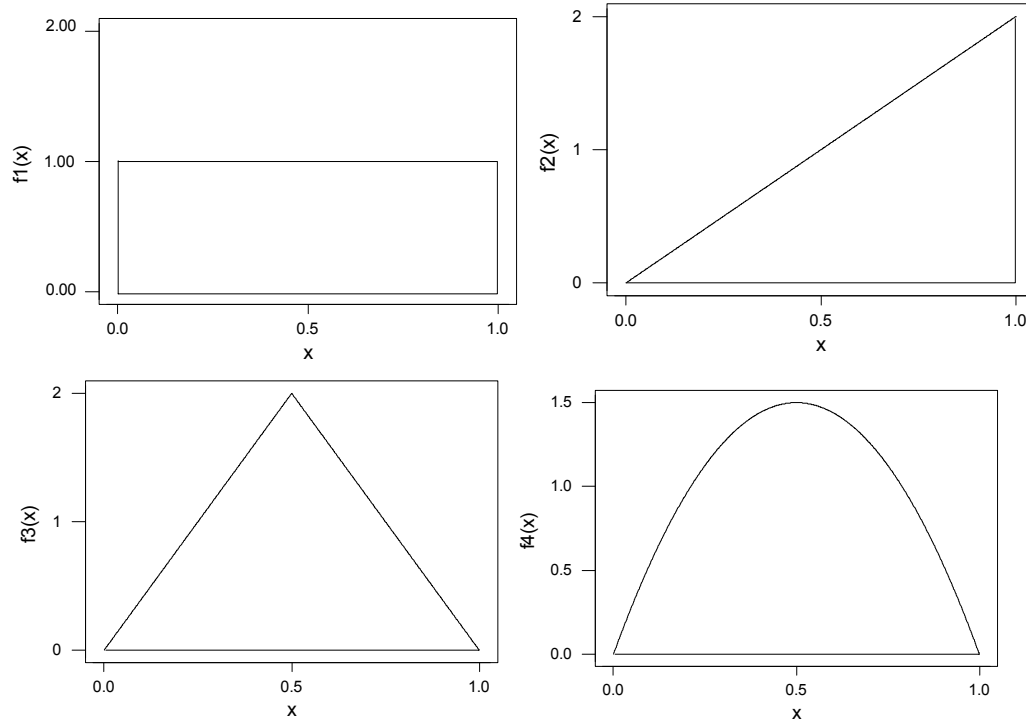


Stat 321 - Day 19
More on Continuous Probability Distributions
(Cumulative Distribution Functions, Expected Values)

Recall that continuous probability distributions are characterized by a probability density function (pdf), with the property that the probability of the random variable falling in some interval equals the area under the pdf over that interval.

Example: Random Lunch Times (cont.)

Recall the example of a businessperson who leaves for lunch at a time between noon and 1:00pm that varies from day to day. Let the random variable X =time (in hours) after noon that the person leaves for lunch. Consider the following four probability density functions for X :



You found that functional expressions for these pdf's are:

$$f_1(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_2(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_3(x) = \begin{cases} 4x & 0 < x < 1/2 \\ 4(1-x) & 1/2 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_4(x) = \begin{cases} 6x(1-x) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Another way to characterize a continuous probability distribution is with a cumulative distribution function (cdf). This function is defined just as it was with discrete distributions: $F(x)=P(X \leq x)$. Because probabilities correspond to integrals with continuous probability distributions, this function can also be written as $F(x) = \int_{-\infty}^x f(t)dt$, where $f(x)$ is the pdf and t is a dummy variable of integration.

- (a) Determine the cdf for each of these pdf's. Also draw sketches of the cdf's. [Hints: In all cases, be sure to specify what the cdf outputs for all real number inputs. Use geometric arguments to help make sense of the integrals, especially in case #3.]

case #1:

Sketch:

case #2:

Sketch:

case #3:

Sketch:

case #4:

Sketch:

Now consider a fifth probability distribution, whose cdf is given by $F_5(x) = \begin{cases} 0 & x \leq 0 \\ x^4 & 0 < x < 1. \\ 1 & x \geq 1 \end{cases}$.

(b) Sketch this cdf.

(c) Determine $P(X < 1/4)$, $P(X > 3/4)$, and $P(1/3 < X < 2/3)$ directly from this cdf.

$$P(X < 1/4) =$$

$$P(X > 3/4) =$$

$$P(1/3 < X < 2/3) =$$

(d) Determine the pdf, and sketch it.

You should have found that the cdf of a continuous distribution can lead directly to probability calculations:

- $P(X \leq b) = F(b)$
- $P(X > a) = 1 - F(a)$
- $P(a < X \leq b) = F(b) - F(a)$.

You should also have noted that the pdf may be found from the cdf from $f(x) = F'(x)$.

(e) Explain why, with a continuous random variable, the following are also true:

- $P(X < b) = F(b)$
- $P(X \geq a) = 1 - F(a)$
- $P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b) = F(b) - F(a)$.

As with the discrete case, one can calculate the expected (mean) value and variance of a continuous random variable: $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$, $V(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2$, where one find the expected value of a function of the random variable by the expression:

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx.$$

(f) For the five probability distributions that you have analyzed above, make a guess for the mean (expected) value of each. Record your guesses in the table below. Then use calculus to determine these expected values, record them in the table as well, and comment on how well your guesses did.

distribution	1	2	3	4	5
guess for mean					
expected value					

The interpretation of expected (or mean) value is the same for continuous distributions as for discrete ones: the long-term average value that would result from repeating the process over and over.

(g) Judging from the variation in the graphs of the pdf's, make a guess for the relative ranking of their standard deviations. Record your guesses (by case number) in the following table. Then calculate their standard deviations, record their values in the table, and comment on how well your guesses did.

	smallest	next smallest	middle	next largest	largest
guess					
std dev value					

The $(100p)$ th *percentile* of a probability distribution is the value (call it k) such that $P(X \leq k) = p$. In particular, the *median* is the 50th percentile, the *lower quartile* is the 25th percentile, and the *upper quartile* is the 75th percentile.

(h) Determine the median, lower quartile, and upper quartile for the fifth probability distribution described above. Comment on how the mean and median compare.

Median:

Lower quartile:

Upper quartile: