18.01A Topic 15: Continuous random variables, standard deviation.
Read: SN: P sections 2,3

Read the supplementary notes

Continuous random variables
Example: Let X be the waiting time between requests at a telephone switch. Unlike our previous examples X can take any positive value –it is continuous.

This is nice, we can apply the calculus we know. For instance, instead of sums we get to take integrals.

Definition: A **continuous random variable** X is a variable x together with a **probability density function** $f(x)$ such that
i) X takes values in (x_1, x_2) (here x_1 can be $-\infty$ and x_2 can be ∞).
ii) $f(x) \geq 0$.
iii) $\displaystyle\int_{x_1}^{x_2} f(x)\,dx = 1$.
iv) $P(a \leq X \leq b) = \displaystyle\int_a^b f(x)\,dx$.

Exponential distribution with parameter m
i) Values in $[0, \infty)$.
ii) $f(x) = \dfrac{\mathrm{e}^{-x/m}}{m}$, where $m > 0$.
iii) Exponential distributions are used to model waiting times.
Soon we will define the following for X an exponential R.V. with parameter m):
iv) $E(X) = m, \quad \sigma^2(X) = m^2$ (*expected value and variance*).
v) $F(x) = 1 - \mathrm{e}^{-x/m}$ (*cumulative distribution function*).

It's easy to check that $\displaystyle\int_0^\infty f(x)\,dx = 1$, i.e., $\displaystyle\int_0^\infty \frac{\mathrm{e}^{-x/m}}{m}\,dx = -\mathrm{e}^{-x/m}\Big|_0^\infty = 1$.

Example: Let X be an exponential R.V. with mean $m = 1$. Find $P(0 \leq X \leq 1)$.
<u>**answer:**</u> $P(0 \leq X \leq 1) = \displaystyle\int_0^1 \mathrm{e}^{-x}\,dx = 1 - \mathrm{e}^{-1} = .632$.

Example: Same question if $m = 10$.
<u>**answer:**</u> $P(0 \leq X \leq 1) = \displaystyle\int_0^1 \frac{\mathrm{e}^{-x/10}}{10}\,dx = 1 - \mathrm{e}^{-1/10} = .095$.

What is $P(X \geq 60)$?
<u>**answer:**</u> $P(60 \leq X) = \displaystyle\int_{60}^\infty \frac{\mathrm{e}^{-x/10}}{10}\,dx = \mathrm{e}^{-60/10} = .002$.

(continued)

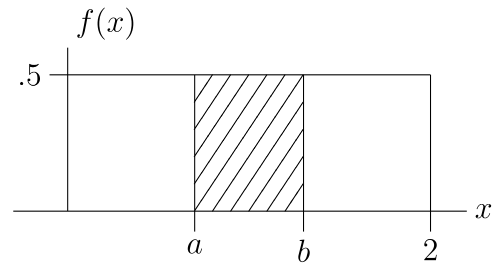
Uniform Distribution on $[x_1, x_2]$ X = the simplest continuous random variable.

i) Values in $[x_1, x_2]$.

ii) $f(x) = \dfrac{1}{x_2 - x_1}$.

iii) $E(X) = \dfrac{x_1 + x_2}{2}$, $\sigma^2(X) = \dfrac{(x_2 - x_1)^2}{2}$.

Check: $\int_{x_1}^{x_2} f(x)\,dx = 1$ (easy).



Uniform density on $[0, 2]$.

Note: shaded area $= P(a \le X \le b) = \dfrac{b-a}{x_2 - x_1}$.

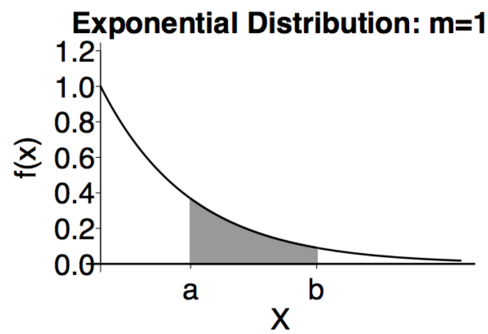
Histograms

Let X be an exponential distribution with mean 10

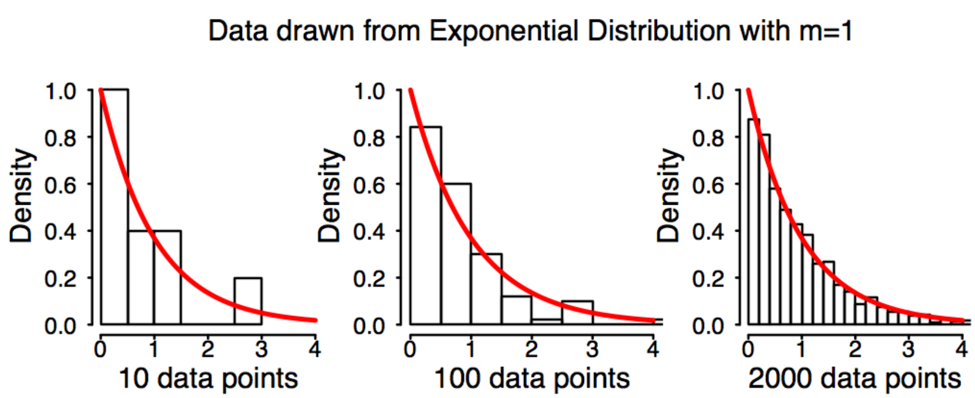
$\Rightarrow$ $f(x) = \dfrac{1}{m}e^{-x/m}$ is its probability density function.

The shaded area $= P(a \le X \le b) = \int_a^b f(x)\,dx$.

Concept question: What is the area under the curve from $x = 0$ to ∞?

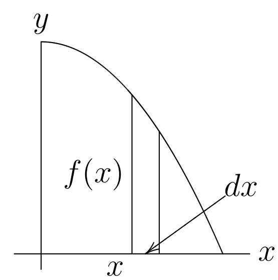


The next picture shows histograms made from data taken (using a statistical package) from an exponential distribution. The histograms are scaled to have total area 1. Notice how well they approximate the density.



(continued)

Expectation: $E(X) := \int_{x_1}^{x_2} x\, f(x)\, dx.$



Why? Slice and sum.

For discrete random variables $E(X) = \sum x_j P(x_j)$.

For the continuous r.v. the sum is replaced by and integral where the value of X in the slice is x and $P(X \text{ is in slice}) = f(x)\, dx$.

The expected value has the same interpretation as before. That is, if you took the average of a large number of independent trials it should be close to the expected value.

Example: X a uniform random variable on $[x_1, x_2]$.

$$\Rightarrow\ E(X) = \frac{1}{x_2 - x_1}\int_{x_1}^{x_2} x\, dx = \frac{1}{x_2 - x_1}\left.\frac{x^2}{2}\right|_{x_1}^{x_2} = \frac{x_2^2 - x_1^2}{2(x_2 - x_1)} = \frac{x_2 + x_1}{2}.$$

Example: X an exponential random variable with parameter m.

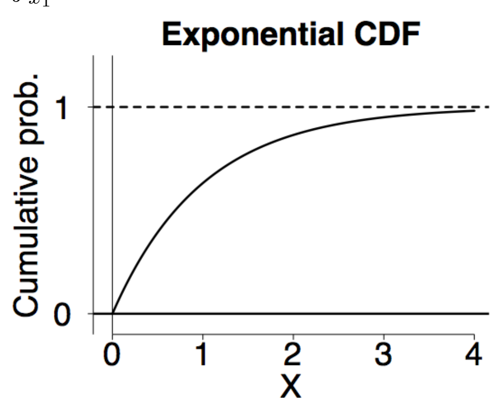
$\Rightarrow\ E(X) = \int_0^\infty x\frac{e^{-x/m}}{m}\, dx = -xe^{-x/m} - me^{-x/m}\Big|_0^\infty = m$. (The integral is done by parts, and we need to know how to take the limit at ∞.)

Cumulative Distribution Function $F(x) := \int_{x_1}^{x} f(x)\, dx = P(X \le x)$.

i) $F(x)$ is an antiderivative of $f(x)$.

ii) $F(x)$ is increasing.

iii) $\lim_{x\to\infty} F(x) = 1$.



Example: X an exponential random variable with mean m.

$$F(x) = \int_0^x \frac{e^{-u/m}}{m}\, du = e^{-u/m}\Big|_0^x = 1 - e^{-x/m}.$$

Example: (2.2 in supplementary notes). A radioactive substance emits a *beta*-particle on average every 10 seconds. What's the probability of waiting more than a minute for the next emission?

answer: Radioactive waiting times are modeled by an exponential distribution X with mean m.

Find m: m is the average time $\Rightarrow$ $m = 10$ seconds.

$P(X > 60) = 1 - P(X < 60) = 1 - F(60) = 1 - (1 - e^{-6}) = e^{-6} = .002$. (Very small probability.)

(continued)

Variance and Standard Deviation

If X is a discrete random variable taking values $x_1, x_2, \ldots$ then we define

$m := \sum_i x_i \, P(x_i)$ (*mean or expected value* of X).

$\sigma^2 := \sum_i (x_i - m)^2 \, P(x_i)$ (*variance* of X).

$\sigma := \sqrt{\sigma^2}$ (*standard deviation* of X).

If X is a continuous random variable with range $[x_1, x_2]$ then we define

$m := \int_{x_1}^{x_2} x \, f(x) \, dx$ (*mean or expected value* of X).

$\sigma^2 := \int_{x_1}^{x_2} (x - m)^2 \, f(x) \, dx$ (*variance* of X).

$\sigma := \sqrt{\sigma^2}$ (*standard deviation* of X).

Theorem: We have the formulas $\sigma^2 = \sum_i x_i^2 \, P(x_i) - m^2$ or $\sigma^2 = \int x^2 \, f(x) \, dx - m^2$.

Proof: See the supplementary notes.

Notes:

1. Variance is always non-negative.
2. Variance measures the spread of the distribution around the mean.
3. If X is a Poisson r.v. with mean m then $\sigma(X) = \sqrt{m}$.
4. If X is an exponential r.v. with mean m then $\sigma(X) = m$.

The proofs of (3) and (4) are similar to those used to find $E(X)$.

Standard deviation will be very important when we study normal distributions.