Math 183
Statistical Methods

Eddie Aamari
S.E.W. Assistant Professor

eaamari@ucsd.edu
math.ucsd.edu/~eaamari/
AP&M 5880A

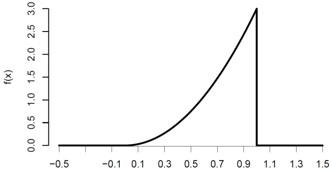
Today: Chapter 3 (continued)

- Common questions with continuous distributions
- Uniform, Exponential and Normal models and their parameters
- Memorylessness of the exponential model
- Finding the area under a Normal curve without finding integrals

From Last Lecture

Suppose the concentration of iodine in a chemical sample is modeled by the density function

$$f(x) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$



Is this a common density function?

What are the common ones?

Common Density Function #1

Sometimes, a finite interval of possibilities are all equally likely:

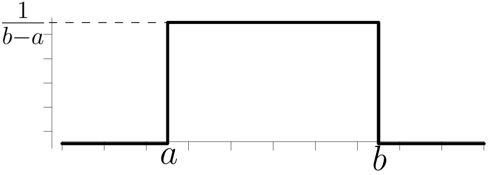
Examples:

- You flick a spinner and measure the angle it makes, in radians, between 0 and 2π.
- You arrive at the bus stop at 8:00am. It is equally likely to show up at any moment in the next 10 minutes.
- The random number generator of R, that outputs random real numbers in the interval $[0, 1]$.

Uniform Distribution

The density of the
uniform distribution
over $[a, b]$ is

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise.} \end{cases}$$



Notation: $X = Unif(a, b)$.

Since the density function is horizontal, it is easy to find areas without the need for calculus.

Intuitively, we would expect $E(X) = \dfrac{a+b}{2}$.

Uniform Distribution: Parameters

If $X = Unif(a, b)$,

$$
\begin{aligned}
E(X) = \int_{-\infty}^{\infty} xf(x)dx &= \frac{1}{b-a}\int_a^b xdx \\
&= \frac{1}{b-a}\left(\frac{x^2}{2}\right)\Bigg|_a^b \\
&= \frac{1}{b-a}\left(\frac{b^2}{2} - \frac{a^2}{2}\right) \\
&= \frac{(b+a)(b-a)}{2(b-a)} = \frac{a+b}{2}.
\end{aligned}
$$

$$Var(X) = \frac{(b-a)^2}{12} \qquad\qquad SD(X) = \frac{b-a}{\sqrt{12}}$$

(Try deriving these yourself!)

Common Density Function #2

My favorite soccer team scores an average of 2.5 goals per game. This is a rate: 2.5 goals/game, or 2.5 goals/90 minutes.

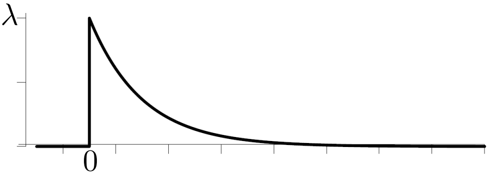
- What is the probability of an 8-goals game?
  $\rightarrow$ This asks about a discrete idea, so use Poisson distribution
- How long do we have to wait, on average, for that first goal?
  $\rightarrow$ This asks about a continuous idea, so use exponential distribution

The exponential distribution models the amount of time we have to wait for an event that occurs with frequency λ (units/time interval)

Exponential Distribution

The density of the **exponential distribution** is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$



where $\lambda > 0$ describes a time-based rate.

Notation: $X = Exp(\lambda)$.

Is the area under $f(x)$ really equal to 1?

$$\int_0^\infty \lambda e^{-\lambda x} dx = \lim_{n\to\infty} \int_0^n \lambda e^{-\lambda x} dx = \lim_{n\to\infty} \left(-e^{-\lambda x}\right)\Big|_0^n$$

$$= \lim_{n\to\infty} \left(-e^{-\lambda n} - (-1)\right) = 1$$

Exponential Distribution: Memorylessness

Let $X = Exp(\lambda)$. X represents how long we will have to wait before an event occurs (where the event occurs with rate λ).

Example: A politician tells 11 lies per week. If we set $X = Exp(11/7)$, then X is a variable that models how long (in days) we must wait before hearing a lie.

The exponential distribution is said to be memoryless. This means:

$$P(X \geq s + t | X \geq s) = P(X \geq t).$$

- $P(X \geq s + t | X \geq s)$: the probability that you will have to wait extra time t or more for something to occur, given that you have already waited time s.
- $P(X \geq t)$: the probability you will have to wait time t or more for something to occur.

Memorylessness: Example

A politician tells 11 lies per week. If we set $X = Exp(11/7)$, then X is a variable that models how long (in days) we must wait before hearing a lie.
Memorylessness states that

$$P(X \geq 3 + 1 | X \geq 3) = P(X \geq 1).$$

That is to say,

> the probability we can make it a whole day without a lie given that we've already made it 3 days without a lie

is the same than

> the probability we can make it a whole day without a lie.

Moral: The time for something to happen is unaffected by how much time has already passed.

Remark: The exponential distributions $Exp(\lambda)$ are the only distributions having this property.

Exponential Model: Example

Dishwashers tend to break down once every 5 years. If the length of time until a breakdown follows an exponential distribution, what is the probability a dishwasher lasts at least 8 years with no breakdown?

Decide on λ and consider units: $\lambda = \dfrac{1 \text{ breakdown}}{5 \text{ years}} = \dfrac{1}{5}$.

Define your model: Let $X = Exp(1/5)$ be the amount of time (in years) until a breakdown.

Write your question in notation: $P(X \geq 8) =?$.

Exponential Distribution: Example

Do the math:

$$
\begin{aligned}
P(X \geq 8) &= \int_8^\infty \frac{1}{5} e^{-x/5} dx \\
&= \lim_{n\to\infty} \int_8^n \frac{1}{5} e^{-x/5} dx \\
&= \lim_{n\to\infty} (-e^{-x/5}) \Big|_8^n \\
&= \lim_{n\to\infty} \left(-e^{-n/5} - (-e^{8/5})\right) \\
&\boxed{= e^{-8/5} \simeq 0.202.}
\end{aligned}
$$

Be sure to:

- Cut out spots where the density has height 0 (this did not happen here)
- Change infinite bounds to variables and use limits to head there
- Line up equals sign if you work vertically
- Box your answer and give an exact answer and approximation

Exponential Distribution: Parameters

How long do we expect, on average, a dishwasher to go before the first breakdown? Is there much variation in that amount of time?

If $X = Exp(\lambda)$, then

$$E(X) = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2} \qquad SD(X) = \frac{1}{\lambda}$$

(As with the Poisson distribution, this answer is very logical since we used an average rate as the basis for the model.)

In our dishwasher example,

$$E(X) = \frac{1}{1/5} = 5 \text{ years} \qquad SD(X) = \frac{1}{1/5} = 5 \text{ years}$$

Best Dishwasher Ever

You have the best dishwasher ever. It hasn't broken down in 30 years. What is the probability you get at least 8 more years before the first breakdown?

We want $P(X \geq 30 + 8 | X \geq 30)$.

By the memoryless property,

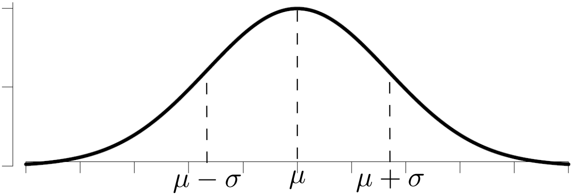
$$P(X \geq 30 + 8 | X \geq 30) = P(X \geq 8) = \boxed{e^{-8/5} \simeq 0.202.}.$$

The Normal Distribution

The density function of the **normal distribution** is:

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

where μ is the mean, and σ is the standard deviation.



Notation: $X = N(\mu, \sigma)$.

Other name: Gaussian distribution

Importance of the Normal Distribution

The Normal model is the most important continuous random variable in all of modern statistics.

Roughly speaking, this comes from the fact that any time some quantity is the combination of many independent factors, then this quantity will follow a normal model.

Examples:

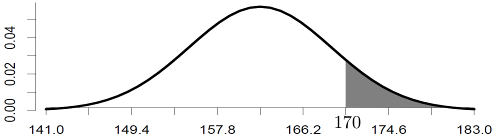
- Human heights in the US (Female: $\mu = 162$ cm, $\sigma = 7$ cm)
- Diastolic blood pressure ($\mu = 77$ mm Hg, $\sigma = 5.5$ mm Hg)
- IQ scores ($\mu = 100$, $\sigma = 15$)

Normal Distribution: Example

Suppose the height of US women is normally distributed with a mean $\mu = 162$ cm and standard deviation $\sigma = 7$ cm.
What is the probability the next woman you see has a height over 170 cm?

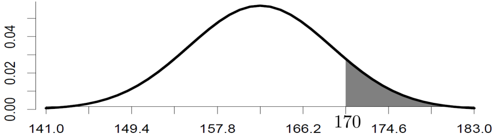
Let $X = N(162, 7)$. We want

$$P(X \geq 180) = \int_{170}^{\infty} \frac{1}{7\sqrt{2\pi}} e^{-\frac{(x-162)^2}{2\times 7^2}} dx$$
$$= ...$$



That's a hard integral! This is the first example of a distribution where probabilities cannot be found by hand.

Technology and the z-table



R provides us with an answer with the function `pnorm`.

```
> pnorm(170, mean = 162, sd = 7, lower.tail = F)
[1] 0.126549
```

$$P(X \geq 170) \simeq 12.65\%.$$

Another way to find the answer is the so-called z-table.

- It is a way to look up areas under part of a normal curve.
- Trouble: we don't want a different table for every possible normal curve (μ and σ can take infinitely many values)
  We need a way to convert any situation modeled by a normal curve into some standard setup.

More on this next class!