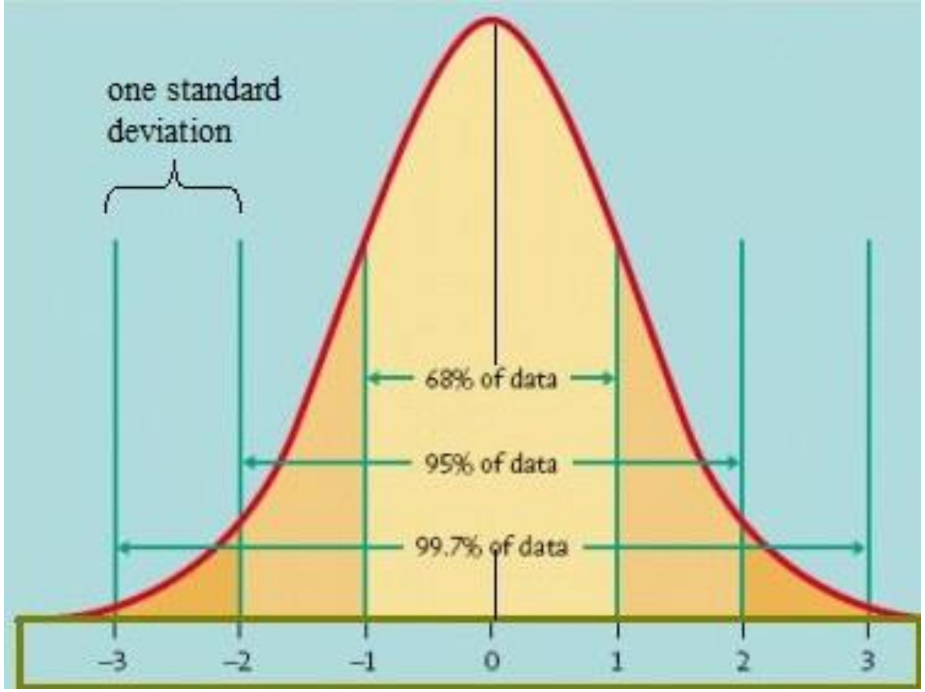
Stat 110

Unit 5: Continuous Random Variables

Ch. 5 in the text



Unit 5 Outline

- Probability Density Functions
- Uniform Distribution
- Universality of the Uniform
- Normal Distribution
- Exponential Distribution
- Poisson Processes

Definition: Continuous r.v.

- A r.v. has a ***continuous distribution*** if its CDF is differentiable. We also allow there to be endpoints (or finitely many points) where the CDF is continuous but not differentiable, as long as the CDF is differentiable everywhere else.
- For a continuous r.v. X with CDF F, the ***probability density function*** (PDF) of X is the derivative, f, of the CDF, given by $f(x) = F'(x)$: The support of X, and of its distribution, is the set of all x where $f(x) > 0$.

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

- So how do we calculate probabilities for a continuous random variable?

Probabilities for a Continuous r.v.

- So how can we calculate the probability that a continuous r.v. *X* falls into an interval (*a*,*b*)? Or [*a*,*b*)? (*a*,*b*]? [*a*,*b*]?
- By the definition of CDF and the fundamental theorem of calculus,

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(x)dx$$

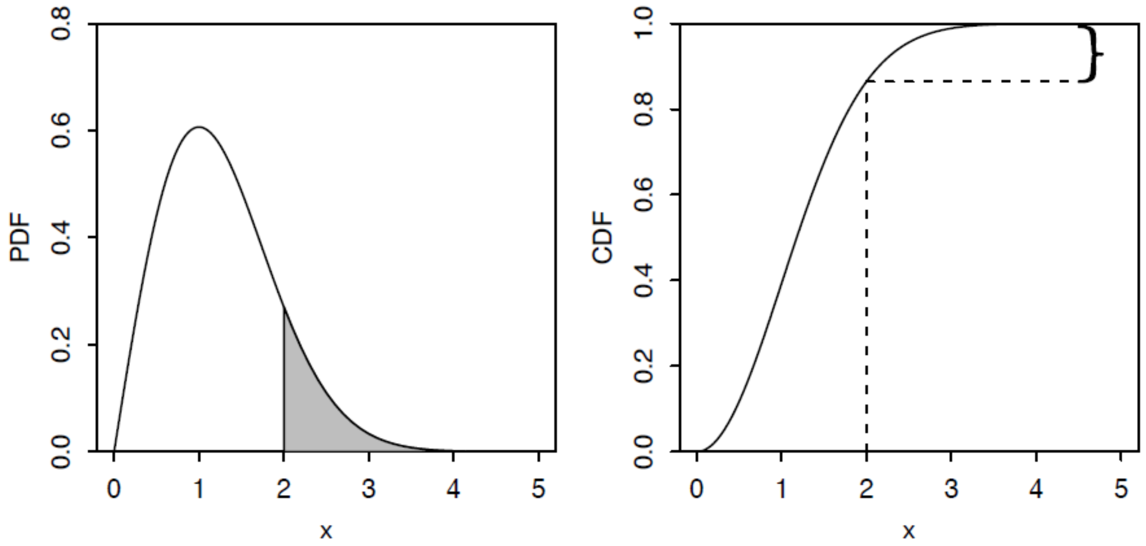
- Translation: to get a desired probability, integrate the PDF over the appropriate range.
- So how is probability represented geometrically?
- Note: We can be carefree about including or excluding endpoints as above for continuous r.v.s (why?), but we must not be careless about this for discrete r.v.s.

Continuous r.v. Example

- Let X have a *Rayleigh* distribution, which has the CDF:

$$F(x) = 1 - e^{-x^2/2}, \quad x > 0$$

- Calculate the PDF of X.
- Find $P(X > 2)$.



Interpretation of a PDF

- So what does $f(x)$ represent anyway?
- We know $f(x)$ is not a probability; for example, we could have $f(3) > 1$, and we know $P(X = 3) = 0$. But thinking about the probability of X being very close to 3 gives us a way to interpret $f(3)$.
- In general, we can think of $f(x)dx$ as the probability of X being in an infinitesimally small interval containing x, of length dx.
- Specifically, the probability of X being in a tiny interval of length ε, centered at 3, will essentially be $f(3)\varepsilon$:

$$P(3-\varepsilon/2 < X < 3+\varepsilon/2) = \int_{3-\varepsilon/2}^{3+\varepsilon/2} f(x)dx \approx f(3)\varepsilon$$

- $f(x)$ is the **density** of the distribution near a specific value x.
- What are the units on $f(x)$?

Expected value of a continuous r.v.

- The ***expected value*** (also called the ***expectation*** or ***mean***, μ) of a continuous r.v. X with PDF f is

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

- The integral is taken over the entire real line, but if the support of X is not the entire real line, we can just integrate over the support.

LOTUS and Variance of a continuous r.v.

- **LOTUS** applies to continuous r.v.s as well. That is:

$$E\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

- The ***Variance*** (σ^2) of a continuous r.v. X with PDF f is

$$\mathrm{Var}(X) = E\left[(X - \mu_X)^2\right] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

- Just like for discrete r.v.s, this is sometimes easier to calculate by using:

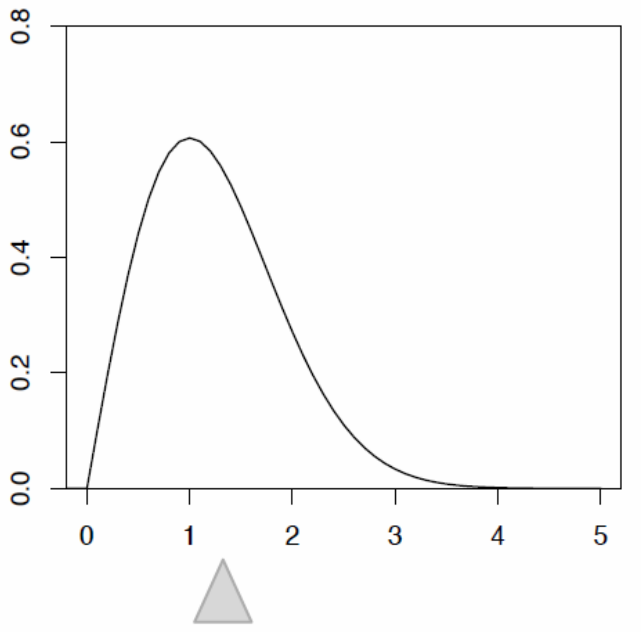
$$\mathrm{Var}(X) = E\left(X^2\right) - \mu_X^2$$

Examples: mean of a continuous r.v.

- Let X ~ Rayleigh. Find $E(X)$. Yay, integration by parts!

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx$$

- What is the interpretation of this value?
- Calculate Var(X).



Unit 5 Outline

- Probability Density Functions
- Uniform Distribution
- Universality of the Uniform
- Normal Distribution
- Exponential Distribution
- Poisson Processes

Story of the Uniform Distribution

- Consider a completely random number (with real value) between the values a and b, each with equal likelihood.
- Let the r.v. X be the value of this completely random number on the interval (a,b).
- Then X has the Uniform distribution with parameters a and b; we denote this by

  $$X \sim \text{Unif}(a,b)$$
- What is X's distribution? That is, what is the probability density function for X? Don't forget to mention X's support.

Uniform Distribution Definition

- If X ~ Unif(a, b), then the PDF of X is:

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

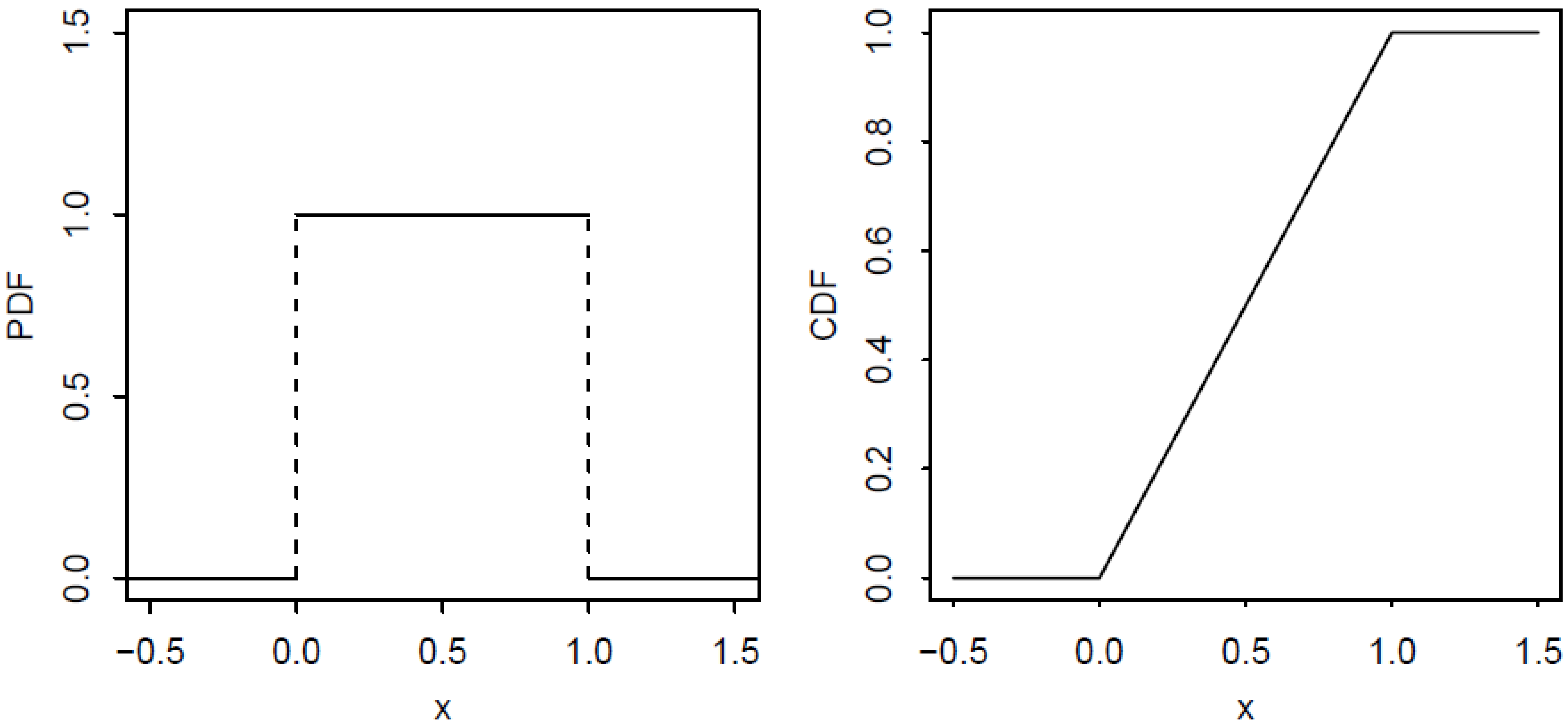
- Or alternatively written:

$$f(x) = \frac{1}{b-a},\ a < x < b$$

- Why is this a valid PMF?
- The most common uniform r.v.? X ~ Unif(0,1). This is sometimes called the *standard uniform*.

Plot of a standard Uniform PDF and CDF

[Unif(0,1)]



The Uniform Distribution: Mean and Variance

- Let $X \sim \text{Unif}(a,b)$.
- Intuitively, what should be the mean of X? What about it's variance?
- Find $E(X)$ and $\text{Var}(X)$.

- $E(X) = (a+b)/2$
- $\text{Var}(X) = (b-a)^2/12$

Unit 5 Outline

- Probability Density Functions
- Uniform Distribution
- Universality of the Uniform
- Normal Distribution
- Exponential Distribution
- Poisson Processes

Universality of the Uniform

- The Unif(0,1) distribution has a remarkable property: given a Unif(0,1) r.v., we can construct **any continuous distribution**.
- Let F be a CDF which is a continuous function and strictly increasing on the support of the distribution. This ensures that the inverse function F^{-1} exists, as a function from (0, 1) to $\mathbb{R}$. We then have the following results.

1. Let $U \sim$ Unif(0,1) and $X = F^{-1}(U)$. Then X is an r.v. with CDF F.
2. Let X be an r.v. with CDF F. Then $F(X) \sim$ Unif(0, 1).

- Don't get confused by what the 2nd part is saying:

Universality of the Uniform (cont.)

- The first part of the theorem says that if we start with $U \sim$ Unif(0, 1) and a CDF F, then we can create a r.v. whose CDF is F by plugging U into the inverse CDF F^{-1}.
- The second part of the theorem goes in the reverse direction, starting from an r.v. X whose CDF is F and then creating a Unif(0, 1) r.v.
- Be careful with the 2nd part: it would be incorrect to say "$F(X) = P(X \leq X) = 1$". Rather, we should first find an expression for the CDF as a function of x, then replace x with X to obtain a random variable. For example, if the CDF of X is $F(x) = 1 - e^{-x}$ for $x > 0$, then $F(X) = 1 - e^{-X}$.

Example: Universality of the Rayleigh

- Let X have a *Rayleigh* distribution, which has the CDF:

$$F(x) = 1 - e^{-x^2/2}, \quad x > 0$$

- Calculate the quantile function F^{-1} (inverse of the CDF).

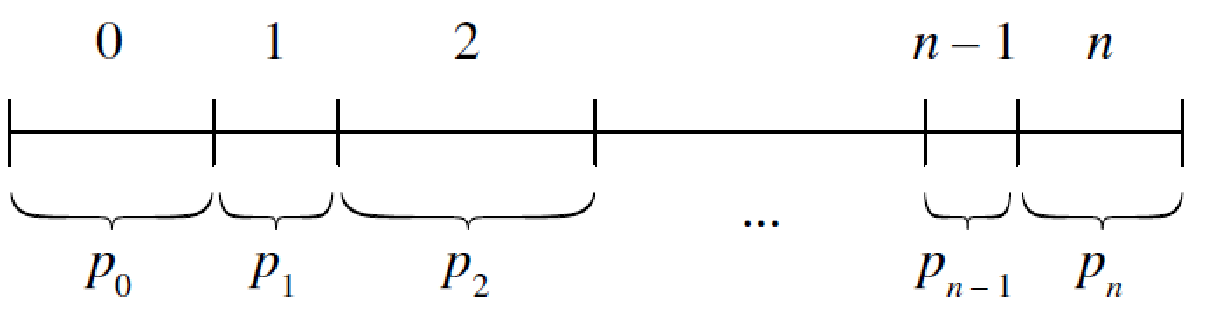
- So if $U \sim \text{Unif}(0,1)$, then

$X = F^{-1}(U) = \sqrt{-2\ln(1-U)} \sim$ Rayleigh.

- What's the support?

Universality discrete r.v.s

- How does the universality of the Uniform hold for discrete random variables?
- The CDF F of a discrete r.v. has jumps and flat regions.
- So a closed form of F^{-1} does not exist.
- But we can still construct any discrete distribution we want from the Uniform. How?
- Given a PMF, chop up the interval (0, 1) into pieces, with lengths given by the PMF values.
- It's best illustrated with a picture:



Why do we care about Universality?

- The Universality of the Uniform is VERY useful when running simulations.
- If we have a method to generate realizations from a Uniform distribution, then we can easily create realizations from any specific distribution.
- Key: we need to know the inverse CDF function (quantile function, F^{-1}) to make this easy.
- Let's use an analogy:
  - Random variables are like houses
  - Distributions are like blueprints
  - Universality gives us a simple rule for remodeling the Uniform house into a house with any other blueprint ☺

Unit 5 Outline

- Probability Density Functions
- Uniform Distribution
- Universality of the Uniform
- Normal Distribution
- Exponential Distribution
- Poisson Processes

Story of the Normal Distribution

- Probably the most famous distribution of all is the Normal distribution (sometimes called the Gaussian).
- It shows up a lot in statistics because of the central limit theorem (which we define more deeply at a later time), which says that under very weak assumptions, the sum (or average) of a large number of i.i.d. random variables has an approximately Normal distribution, regardless of the distribution of the individual r.v.s.
- Example: Heights of individuals. Standardized test scores. Why can these reasonably be assumed to be normally distributed?

Standard Normal Distribution Definition

- We will start off with a special case of the Normal distribution, called the standard Normal, as it will allow to build some properties of Normal r.v.s easily.
- A continuous r.v. Z is said to have the ***standard Normal distribution*** if its PDF φ is given by:

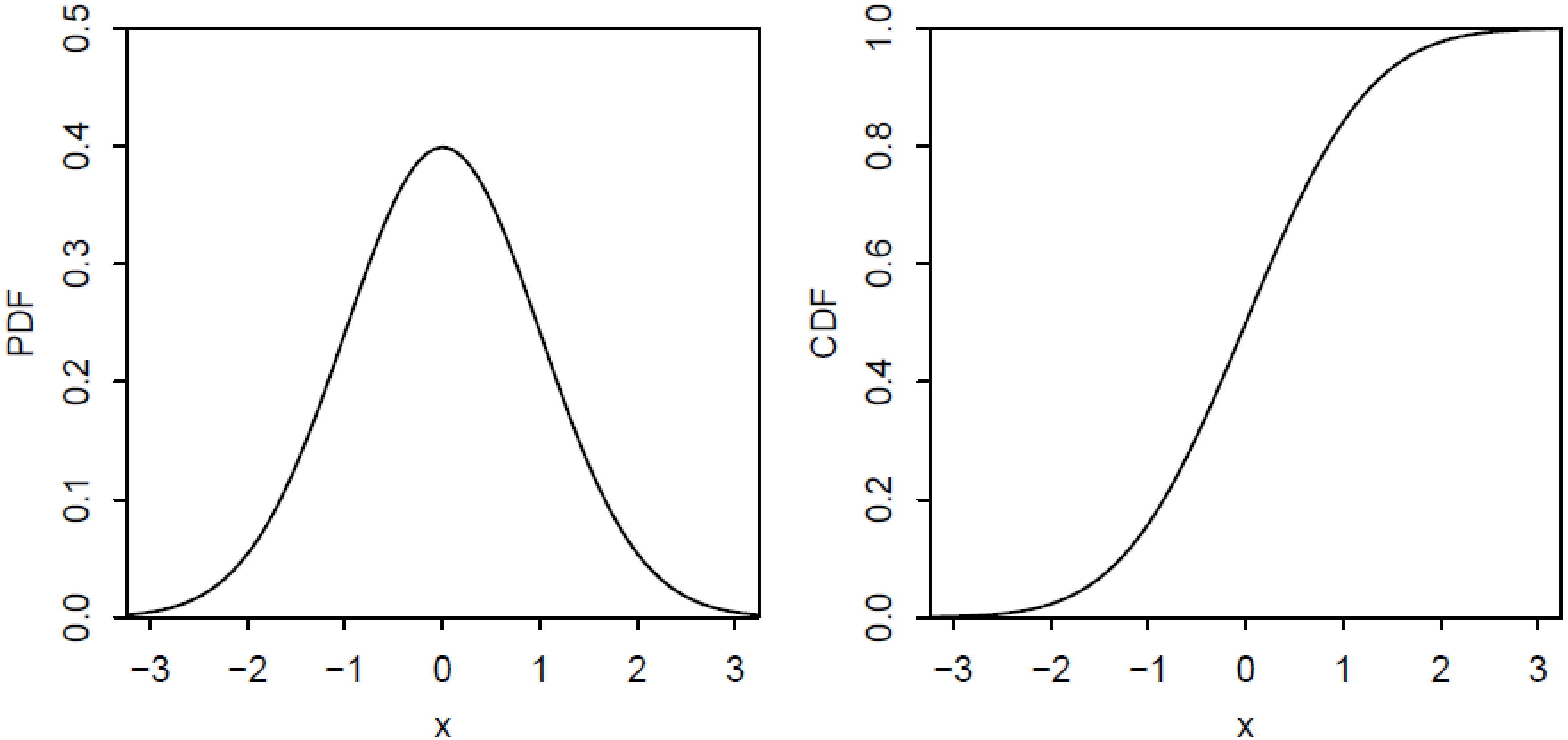
$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \; -\infty < x < \infty$$

- The standard normal CDF Φ is the accumulated area under the PDF:

$$\Phi(z) = \int_{-\infty}^{z} \varphi(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Plot of a standard Normal PDF and CDF

[Normal(0,1)]



Some Properties of the standard Normal

- Symmetry of PDF: $\varphi(z) = \varphi(-z)$.
- Symmetry of tail probabilities: $\Phi(z) = 1 - \Phi(-z)$
- Symmetry of Z and $-Z$.
- We'll prove 3 facts about the standard Normal, and that will allow us to handle the general form of the Normal distribution:
  - φ is a valid PDF.
  - $E(Z) = 0$.
  - $\mathrm{Var}(Z) = 1$.

φ is a valid PDF

- To show φ is a valid PDF, we actually use a strange trick: we multiply the integral twice, and then convert to polar coordinates (which explains the $1/\sqrt{2\pi}$):

$$\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz\right)\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz\right)$$

$$= \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx\right)\left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy\right)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\left(x^2+y^2\right)/2} dxdy$$

$$= \frac{1}{2\pi} \int_{0}^{2\pi} \int_{0}^{\infty} e^{-r^2/2} rdrd\theta$$

φ is a valid PDF (cont.)

- Next step: substitute $u = r^2/2$, $du = rdr$. Which results in:

$$\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{-r^2/2} r\,dr\,d\theta = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{-u}\,du\,d\theta$$

$$= \frac{1}{2\pi}\int_0^{2\pi} 1\,d\theta = \frac{2\pi}{2\pi} = 1$$

- Therefore, since the product of two of the whole probabilities is 1, then each whole probability must be 1.
- Neat trick, huh?

$E(Z) = 0$

- Now we just need to show $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$. Luckily these aren't nearly as tricky.
- Intuitively, why must the $E(Z) = 0$? Math is "easy" too:

$$E(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz$$

- Since $g(z) = z e^{-z^2/2}$ is an odd function, the area from $-\infty$ to 0 cancels with the area from 0 to ∞. Thus, $E(Z) = 0$.

Var(Z) = 1

- Lastly we need to show Var(Z) = 1. Recall:

$$\text{Var}(Z) = E(Z^2) - \mu_Z^2$$

- Since $E(Z) = \mu_Z = 0$, we just need to worry about $E(Z^2)$.

$$E(Z^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz$$

$$= \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} z^2 e^{-z^2/2} dz$$

- This is true since $z^2 e^{-z^2/2}$ is an even function. Now just use integration by parts to show:

$$\text{Var}(Z) = E(Z^2) = 1$$

Normal Distribution Definition

- If X follows a normal distribution, $X \sim N(\mu, \sigma^2)$, then the PDF of X is ($\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}^+$):

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

- Let $Z \sim N(0,1)$, and let $X = \mu + \sigma Z$. Then $X \sim N(\mu, \sigma^2)$.
- Show that the PDF of X is truly what is above.

  *Hint: take derivative of $\Phi[(x- \mu)/\sigma]$.

Mean and Variance of $N(\mu,\sigma^2)$

- Recall, if $Z \sim N(0,1)$ and $X = \mu + \sigma Z$, then $X \sim N(\mu, \sigma^2)$.
- Using the properties of means and variances, find $E(X)$ and $\mathrm{Var}(X)$.

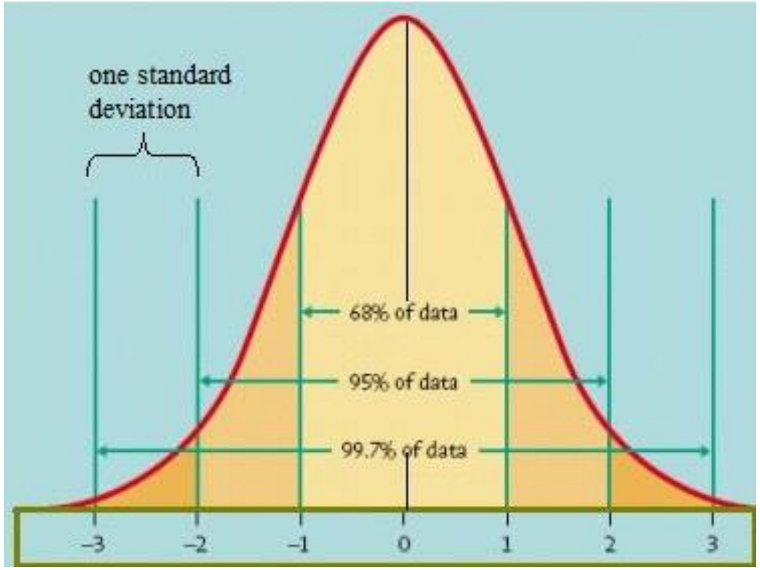
Empirical Rule (68-95-99.7% rule)

- OK, so we saw that taking integrals of the PDF for a Normal distribution is a *little* difficult.
- So we will leave the probabilities in class in terms of Φ.
- Or you could use a computer (or table).
- But it is nice to know the ***empirical rule*** for Normally distributed r.v.s, which states for $X \sim N(\mu, \sigma^2)$:

$$P(|X-\mu| < \sigma) \approx 0.68$$

$$P(|X-\mu| < 2\sigma) \approx 0.95$$

$$P(|X-\mu| < 3\sigma) \approx 0.997$$



Normal Distribution Example (heights of American males)

- Let $X \sim N(\mu = 69, \sigma^2 = 3^2)$. Find the following:
- $P(66 < X < 72)$
- $P(X > 72)$
- $P(60 < X < 72)$
- $P(X > 68)$
- What value of X is needed to be in the top 2.5% of the distribution?

Unit 5 Outline

- Probability Density Functions
- Uniform Distribution
- Universality of the Uniform
- Normal Distribution
- Exponential Distribution
- Poisson Processes

Story of the Exponential Distribution

- Consider you are waiting (in continuous time) until a success for some process/experiment occurs, where λ is the average # successes per unit of time.
- Let X be the amount of time you have to wait until this success arrives.
- Then X has the ***Exponential distribution*** with parameter λ; we denote this by

  $X \sim \text{Expo}(\lambda)$
- The Exponential distribution is the continuous counterpart to the Geometric distribution. How so?
- Example: you are waiting at a bus stop for the next bus to arrive, and the rate of bus arrivals is fixed (λ per unit time).

Exponential Distribution Definition

- If $X \sim \text{Expo}(\lambda)$, then the PDF of X is ($\lambda > 0$):

$$f(x) = \lambda e^{-\lambda x}, \;\; x > 0$$

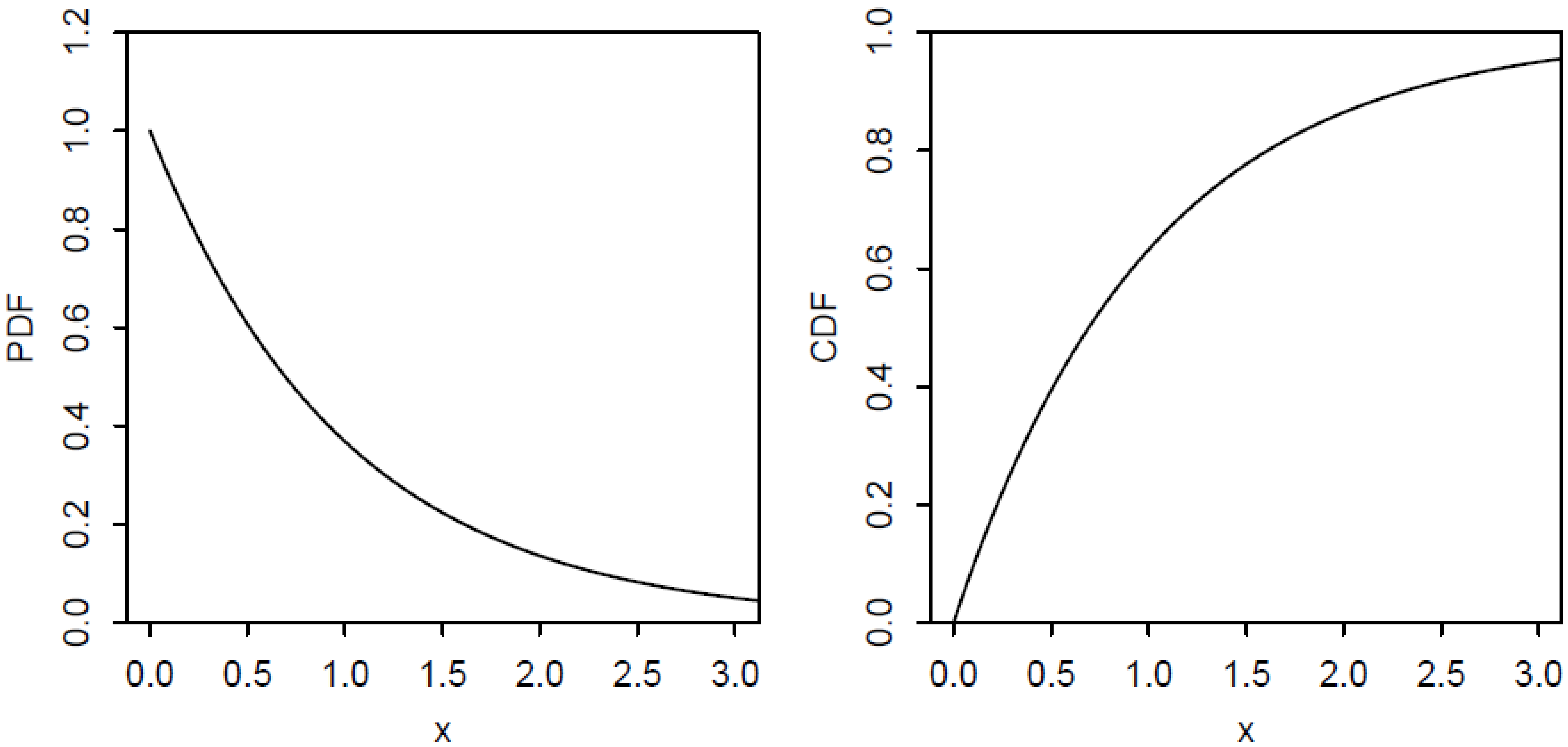
- The corresponding CDF is:

$$F(x) = 1 - e^{-\lambda x}, \;\; x > 0$$

- Why is this a valid PDF?

Plot of an Exponential PDF and CDF

[Expo(1)]



Exponential Distribution Scaling

- Suppose $X \sim \text{Expo}(1)$. Then:

$$Y = \frac{X}{\lambda} \sim \text{Expo}(\lambda)$$

- Intuitively, why does this make sense?
- Using CDFs, show this to be true.

The Exponential Distribution: Mean and Variance

- Let $X \sim \text{Expo}(1)$. Find the mean and variance of X.

  *Note: you will need to use integration by parts.

- Now let $Y = X/\lambda$. What distribution does Y have?
- Use the properties of expectation and variance to find the mean and variance of $Y \sim \text{Expo}(\lambda)$.

Memorylessness property of the Exponential Distribution

- A distribution is said to have *memorylessness property* if a random variable X from that distribution satisfies:

$$P(X \geq s+t \mid X \geq s) = P(X \geq t)$$

- Here s represents the time you've already spent waiting; the definition says that after you've waited s minutes, the probability you'll have to wait another t minutes is exactly the same as the probability of having to wait t minutes with no previous waiting time under your belt.
- This implies that for $X \sim \text{Expo}(\lambda)$:

$$E(X \mid X \geq s) = s + E(X) = s + \frac{1}{\lambda}$$

Memorylessness property of the Exponential Distribution

- Let $X \sim \text{Expo}(\lambda)$. Show that the memorylessness property holds for X.

  *Hint: use the def. of conditional prob. and X's CDF.

- What are the implications of this? Think about waiting at a bus stop if time until next bus arrival is Exponentially distributed. Is this valid in reality?

Exponential Distribution Examples

- OK, so Exponential distributions don't really model human and machine lifetimes. Then why are they useful?
  - Some physical phenomena (radioactive decay) do exhibit the memorylessness property
  - Exponentials are well connected to other distributions (Poisson), and have a shared intuition/story.
  - Exponential distribution can be used as the basis for more flexible distributions (like the *Weibull* dist.) where rate of successes can increase or decrease over time.

Unit 5 Outline

- Probability Density Functions
- Uniform Distribution
- Universality of the Uniform
- Normal Distribution
- Exponential Distribution
- Poisson Processes

Poisson Process Definition

- A process of arrivals in continuous time is called a ***Poisson process*** with rate λ if the following two conditions hold:

1. The number of arrivals that occur in an interval of length t is a Pois(λt) random variable.
2. The numbers of arrivals that occur in disjoint intervals are independent of each other. For example, the numbers of arrivals in the intervals (0; 10); [10; 12); and [15;19) are independent.

Poisson Process Example

- Suppose the emails arrive in your inbox according to a Poisson process with rate parameter λ (0.5 per minute).
- In one hour, *how many* emails, X, will arrive?
  - This follows a Pois(λ) distribution. What is $E(X)$?
- We could also ask: *how long* does it take, T_1, until the first email arrives?
  - What distribution does T_1 follow?
- Saying that the waiting time for the first email is between 0 and t is the same as saying no emails have arrived between 0 and t. So if N_t is the number of emails that arrive at or before time t, then:

  $(T_1 > t)$ is the same event as $(N_t = 0)$

Poisson Process Example

- From last slide: ($T_1 > t$) is the same event as ($N_t = 0$)
- If two events are the same, then they have the same probability (and we know $N_t \sim \text{Pois}(\lambda t)$). So:

$$P(T_1 \geq t) = P(N_t = 0) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}$$

- Thus $P(T_1 < t) = 1 - e^{-\lambda t}$, so $T_1 \sim \text{Expo}(\lambda)$. Wow!
- The time until the 1st arrival in a Poisson process of rate λ has an Exponential distribution with parameter λ. Pretty cool!
- What about $T_2 - T_1$? Since disjoint intervals are independent, the past is irrelevant!
- Note: T_2, which is the sum of two independent Expos is not exponential (it's Gamma distributed).

Minimum of Exponentials is…

- Let $X_1,\ldots, X_n$ be independent with $X_j \sim \text{Expo}(\lambda_j)$. Let $L = \min(X_1,\ldots, X_n)$. Show that : $L \sim \text{Expo}(\lambda_1 + \ldots + \lambda_n)$. *Hint: consider the survival function of L: $S(L) = 1 – F(L)$.

- What does this mean intuitively?

Last Word: what is Normal?