CSE 312
Foundations of Computing II

Lecture 19: Concentration Wrap-Up + Introduction to Continuous Random Variables

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Stefano Tessaro
tessaro@cs.washington.edu

Today – Two things

- Wrap-up concentration / tail inequalities [very quick]
- Introduction to continuous (i.e., non-discrete) random variables

Also: HW5 is online + Naïve Bayes due!

- Extra office hour today with Kushal

Midterm results will come today.

Concentration / tail bounds – A guided tour

https://us.edstem.org/courses/125/discussion/8100

Tail Bounds – Summary

Goal: We need to compute $\mathbb{P}(X > t)$ for $t > \mathbb{E}(X)$

If we know …	… we use …	… to obtain
$\mu = \mathbb{E}(X)$	**Markov**	$\mathbb{P}(X > t) < \frac{\mu}{t}$
$\mu = \mathbb{E}(X)$ and $\sigma^2 = \mathrm{Var}(X)$	**Chebyshev**	$\mathbb{P}(X > t) < \frac{\mathrm{Var}(X)}{(t-\mu)^2}$
$X = X_1 + \cdots + X_n$, sum of indep. RVs in $[0,1]$, $\mu = \mathbb{E}(X)$	**Chernoff**	$\mathbb{P}(X > t) < e^{-\frac{\epsilon^2}{2+\epsilon}\mu}$, where $\epsilon = \frac{t-\mu}{\mu}$

Chernoff usually wins when μ grows as a function of n

Concentration – Summary

Goal: We need to compute $\mathbb{P}(|X-\mu| > \epsilon\mu)$ for $\epsilon > 0$ and $\mu = \mathbb{E}(X)$

If we know …	… we use …	… to obtain
μ only	**Out of luck!**	
$\sigma^2 = \mathrm{Var}(X)$	**Chebyshev**	$\mathbb{P}(\lvert X-\mu\rvert > \epsilon\mu) < \frac{\mathrm{Var}(X)}{\epsilon^2\mu^2}$
$X = X_1 + \cdots + X_n$, sum of indep. RVs in $[0,1]$	**Chernoff**	$\mathbb{P}((\lvert X-\mu\rvert > \epsilon\mu) < 2e^{-\frac{\epsilon^2}{2+\epsilon}\mu}$

Chernoff usually wins when μ grows as a function of n

Sampling Theorem – Recap

- M individuals, a fraction $p \in [0,1]$ is in favor of CSE313
- **Goal:** Produce good estimate $\hat{P}$ of p
- **Idea:**
  - Ask $n < M$ randomly selected individuals whether they want CSE313
  - Responses are Bernoulli $X_1, \dots, X_n$ with parameter p
  - Let $\hat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i \rightarrow \mathbb{E}(\hat{P}) = p$
- **Sampling theorem:** If $n \geq \ln(1/\delta)\frac{2+\theta}{\theta^2}$, then $\mathbb{P}(|\hat{P} - p| \leq \theta) \geq 1 - \delta$
  - θ = how good is the estimate
  - δ = probability we fail to provide good estimate

Sampling Theorem – Proof

$\hat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\mathbb{E}(\hat{P}) = p$ $\mathbb{P}(X_i = 1) = p$

$$\mathbb{P}\left(\left|\hat{P} - p\right| > \theta\right) = \mathbb{P}\left(\left|n\hat{P} - np\right| > n\theta\right)$$
$$= \mathbb{P}(|\textstyle\sum_i^n X_i - np| > n\theta)$$
$$= \mathbb{P}\left(|\textstyle\sum_i^n X_i - np| > np\frac{\theta}{p}\right)$$
$$< 2\exp\left(-\frac{\theta^2/p^2}{2+\theta/p}pn\right)$$
$$= 2\exp\left(-\frac{\theta^2}{2p+\theta}n\right) \leq 2\exp\left(-\frac{\theta^2}{2+\theta}n\right)$$

Need to rephrase in terms of relative error to use Chernoff Bound with $\epsilon = \frac{\theta}{p}$ and $\mu = np$!

Remove dependency on p

Sampling Theorem – Proof (cont'd)

We have proved:

$$\mathbb{P}\left(\left|\hat{P} - p\right| > \theta\right) < 2\exp\left(-\frac{\theta^2}{2+\theta}n\right)$$

We have $2\exp\left(-\frac{\theta^2}{2+\theta}n\right) \leq \delta$ if (and only if)

$$n \geq \ln(1/\delta)\frac{2+\theta}{\theta^2}$$

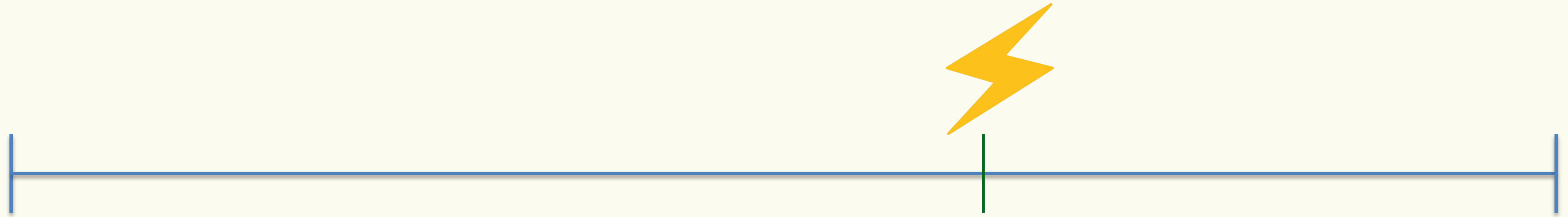
Next – Introduction to Continuous Random Variables

Bottom line: Often we want to model experiments where the outcome is not discrete.

Example – Lightning Strike

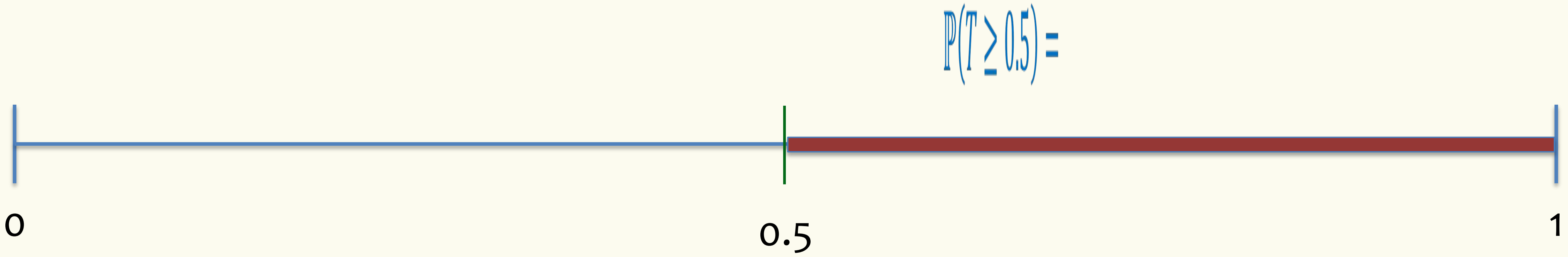
Lightning strikes a pole within a one-minute time frame

- T = time of lightning strike
- Every time within $[0,1]$ is equally likely
  - – Time measured with infinitesimal precision.



Lightning strikes a pole within a one-minute time frame

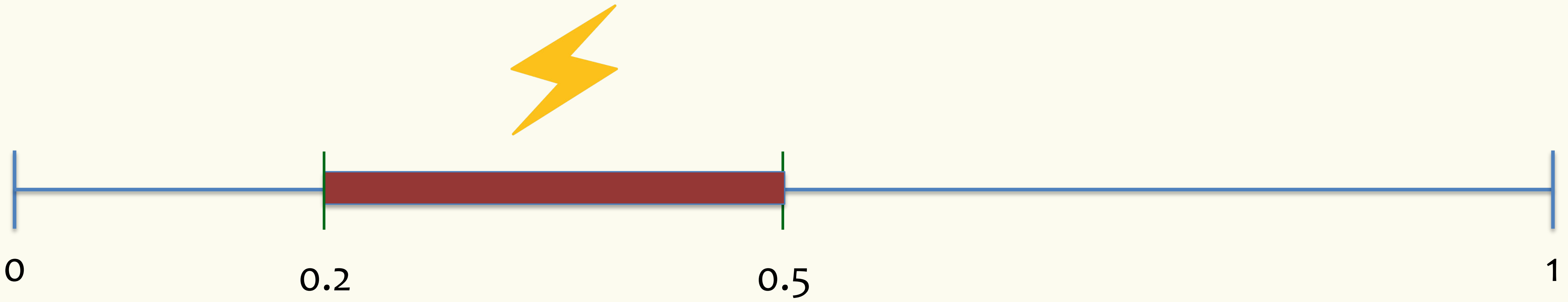
- T = time of lightning strike
- Every point in time within $[0,1]$ is equally likely



$\mathbb{P}(T \geq 0.5) = \frac{1}{2}$

Lightning strikes a pole within a one-minute time frame

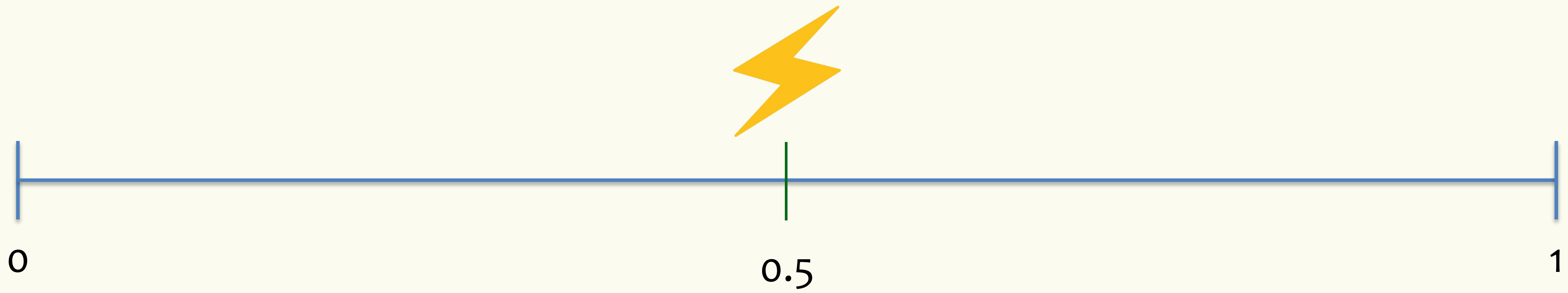
- T = time of lightning strike
- Every point in time within [0,1] is equally likely



$$\mathbb{P}(0.2 \leq T \leq 0.5) = 0.5 - 0.2 = 0.3$$

Lightning strikes a pole within a one-minute time frame

- T = time of lightning strike
- Every point in time within $[0,1]$ is equally likely



$$\mathbb{P}(T = 0.5) = 0$$

Bottom line

- This gives rise to a different type of random variable
- $\mathbb{P}(T = x) = 0$ for all $x \in [0,1]$
- Yet, somehow we want
  - $\mathbb{P}(T \in [0,1]) = 1$
  - $\mathbb{P}(T \in [a,b]) = b - a$
  - …
- How do we model the behavior of T?

Probability Density Function

Definition. A **continuous random variable** X is defined by a **probability density function** (or simply, "density") $f_X: \mathbb{R} \to \mathbb{R}$ such that

- **Non-negativity:** $f_X(x) \geq 0$ for all $x \in \mathbb{R}$
- **Normalization:** $\int_{-\infty}^{+\infty} f_X(x)\,\mathrm{d}x = 1$

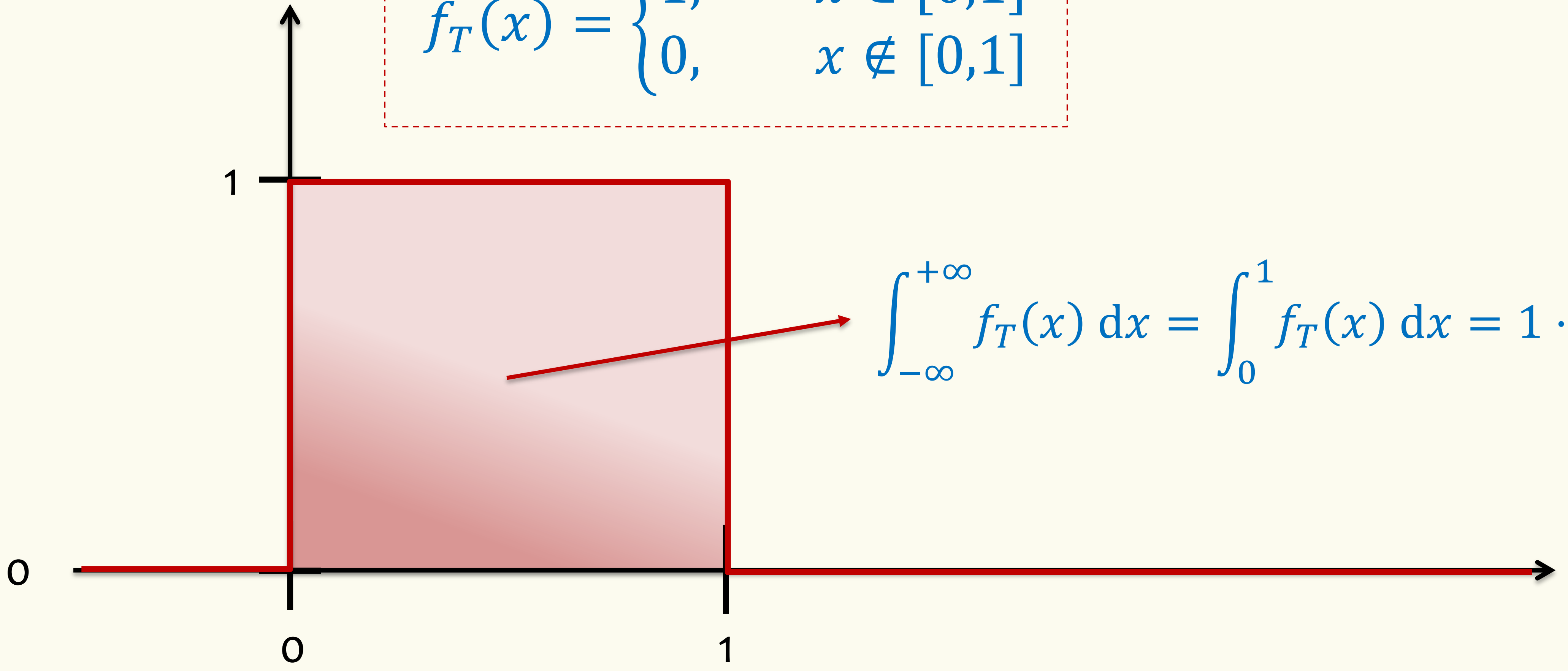
PDF of Uniform RV

Density $\neq$ Probability

$f_T(0.5) = 1 \quad \mathbb{P}(T = 0.5) = 0$

$T \sim \mathrm{Unif}(0,1)$

$$f_T(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}$$



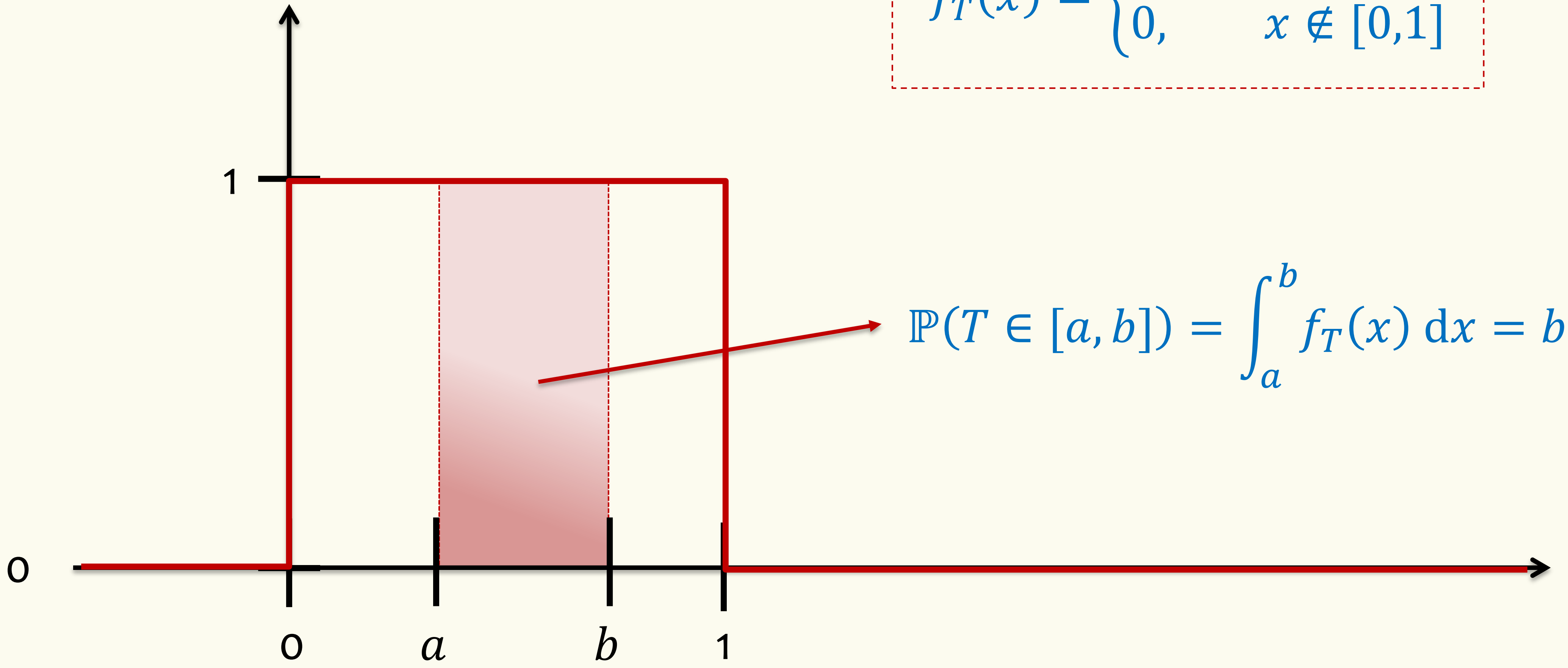
$$\int_{-\infty}^{+\infty} f_T(x)\,\mathrm{d}x = \int_0^1 f_T(x)\,\mathrm{d}x = 1 \cdot 1 = 1$$

Probability of Event

Definition. $\mathbb{P}(X \in S) = \int_S f_X(x)\,\mathrm{d}x$

Example. $T \sim \mathrm{Unif}(0,1)$

$$f_T(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}$$



$$\mathbb{P}(T \in [a,b]) = \int_a^b f_T(x)\,\mathrm{d}x = b - a$$

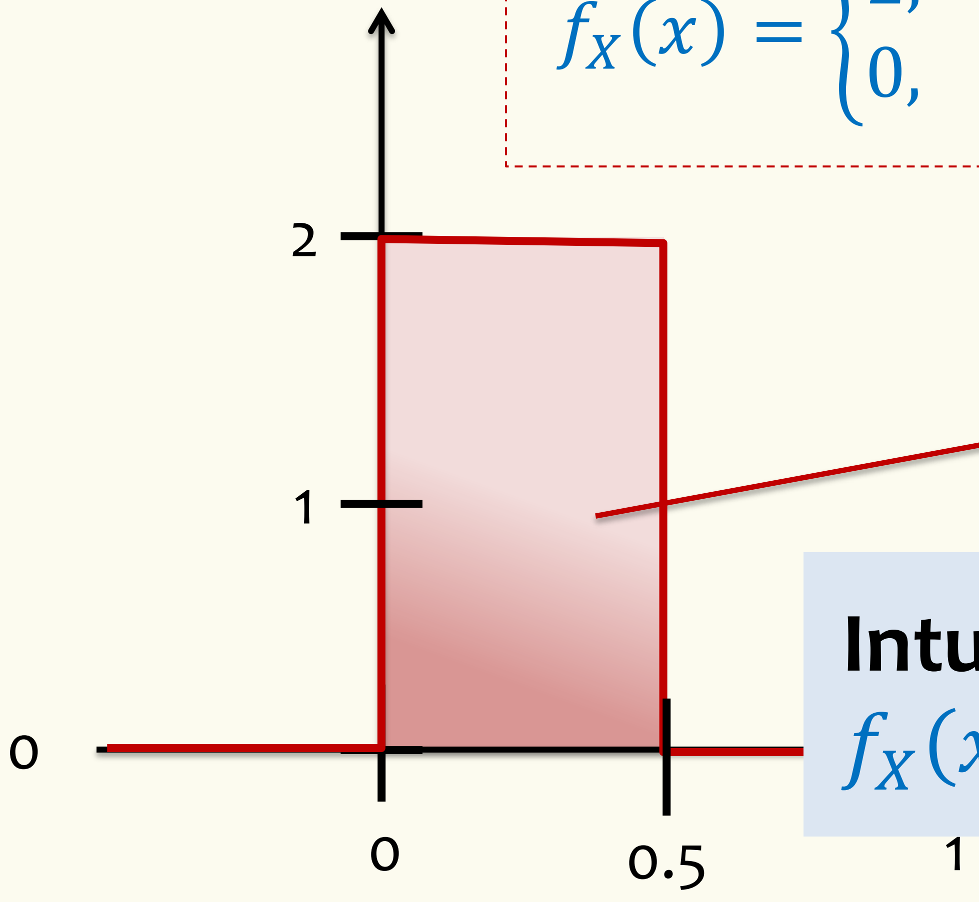
PDF of Uniform RV

Density $\neq$ Probability

$f_X(x) \gg 1$ is possible!

$X \sim \mathrm{Unif}(0,0.5)$

$$f_X(x) = \begin{cases} 2, & x \in [0,0.5] \\ 0, & x \notin [0,0.5] \end{cases}$$



$$\int_{-\infty}^{+\infty} f_X(x)\,\mathrm{d}x = \int_0^1 f_X(x)\,\mathrm{d}x = 2 \cdot 0.5 = 1$$

Intuition: $\mathbb{P}(X \in [x-\epsilon, x+\epsilon]) \approx f_X(x) \cdot \epsilon$ for small ϵ

Cumulative Distribution Function

Definition. The **cumulative distribution function (cdf)** of X is defined as

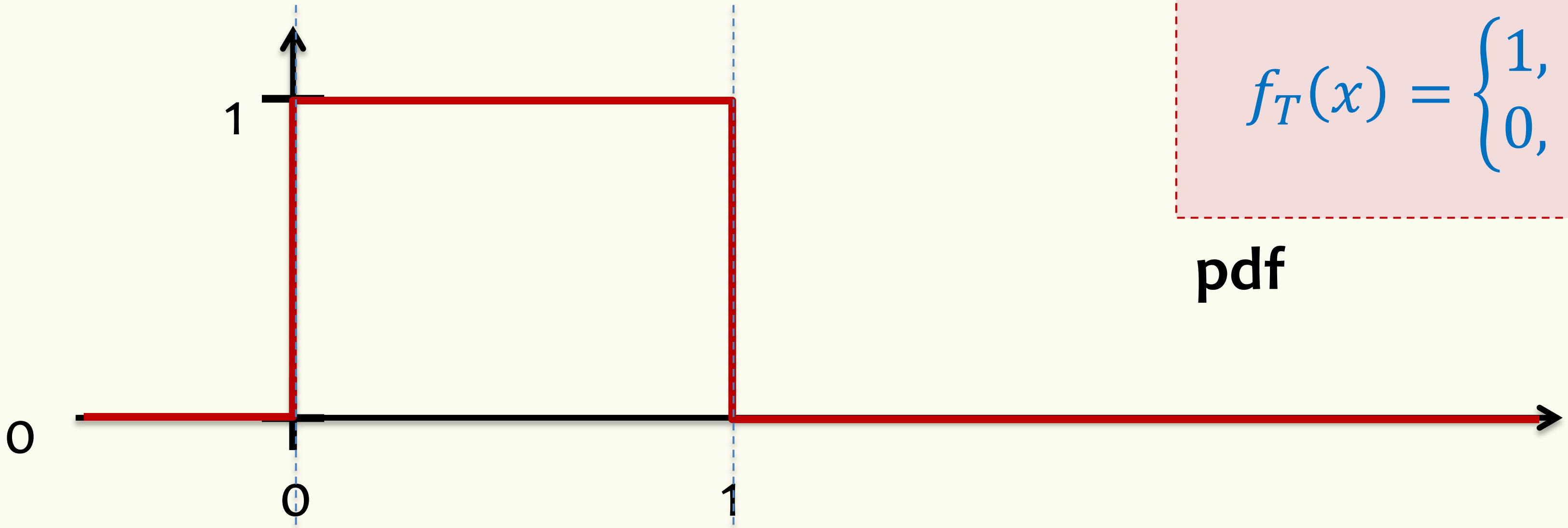
$$F_X(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(x)\,\mathrm{d}x$$

Therefore: $\mathbb{P}(X \in [a, b]) = F(b) - F(a)$

Example. $T \sim \text{Unif}(0,1)$

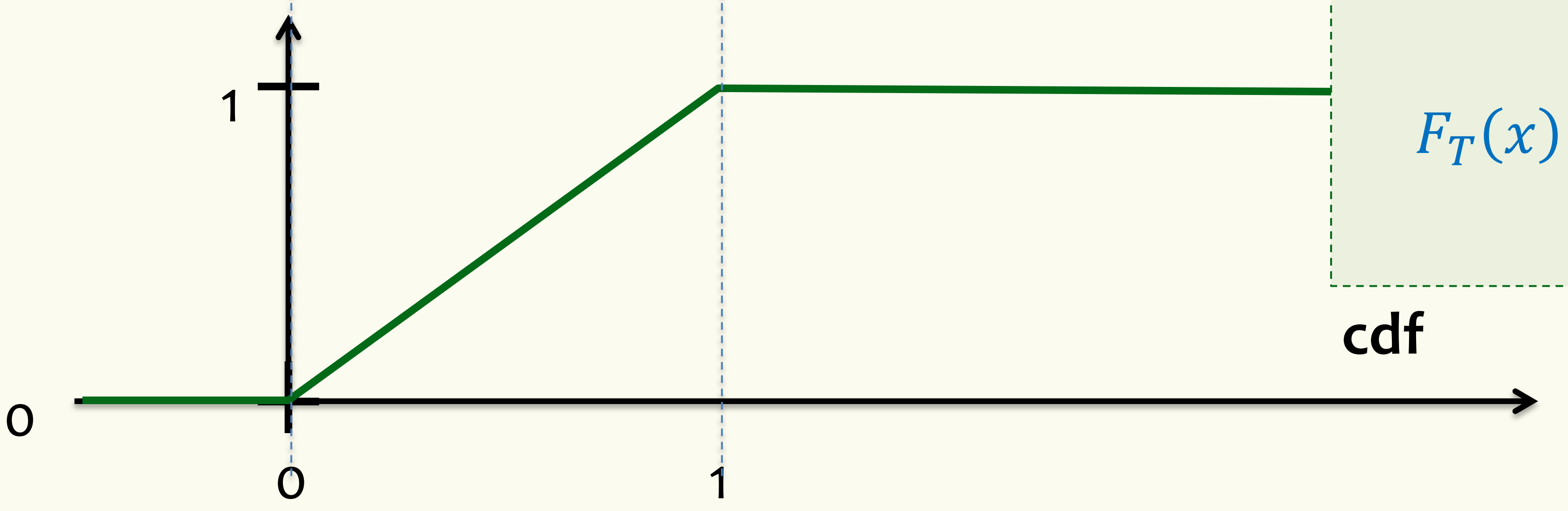
$$f_T(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}$$

pdf



$$F_T(x) = \begin{cases} 0 & x \leq 0 \\ x & 0 \leq x \leq 1 \\ 1 & 1 \leq x \end{cases}$$

cdf



Properties of Random Variables

- Quantities of random variables – $\mathbb{E}(X)$, $\mathrm{Var}(X)$, … – generalize naturally from discrete to continuous RVs
  – Usually have the same properties
- Basic idea: $\sum \to \int$

Expectation of a Continuous RV

Definition. The **expected value** of a continuous RV X is defined as

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} f_X(x) \cdot x \,\mathrm{d}x$$

Fact. $\mathbb{E}(aX + bY + c) = a\mathbb{E}(X) + b\mathbb{E}(Y) + c$

Definition. The **variance** of a continuous RV X is defined as

$$\mathrm{Var}(X) = \int_{-\infty}^{+\infty} f_X(x) \cdot \big(x - \mathbb{E}(X)\big)^2 \,\mathrm{d}x = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

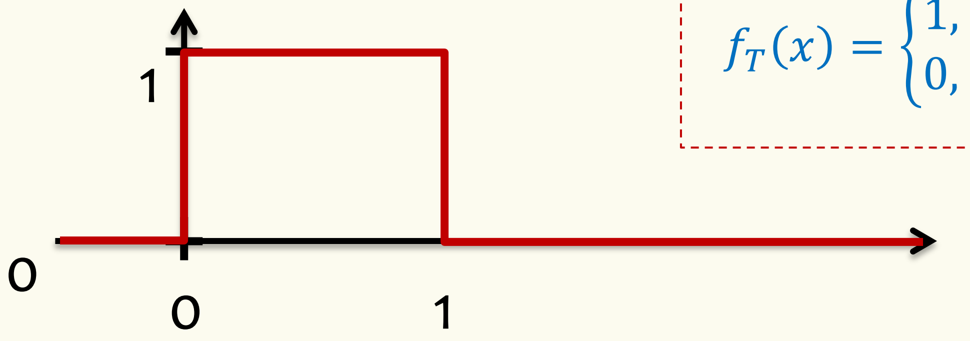
Expectation of a Continuous RV

Definition.

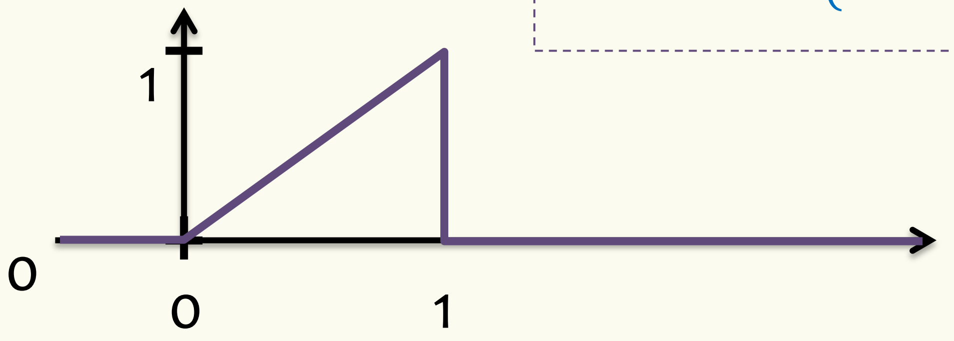
$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} f_X(x) \cdot x \, \mathrm{d}x$$

Example. $T \sim \mathrm{Unif}(0,1)$

$$f_T(x) = \begin{cases} 1, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}$$



$$f_T(x) \cdot x = \begin{cases} x, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}$$



$$\mathbb{E}(T) = \underbrace{\frac{1}{2} 1^2}_{\text{Area of triangle}} = \frac{1}{2}$$