INTRODUCTION TO STATISTICS

LECTURE 3

LAST TIME

- Parameter estimation
- Maximum Likelihood
- Discrete distributions
- Randomized response

TODAY

- Wrap-up randomized response exercise
- Overview of the distributions we’ve seen so far
- More MLE
- Continuous distributions

A bit more
on randomized response

Constrained optimization

REMARKS ON RANDOMIZED RESPONSE

- Ask people two questions:

50% of the cases: **Q1** (embarrassing)

50% of the cases: Q2 (regular, answer is always YES)

REMARKS ON RANDOMIZED RESPONSE

- Ask people two questions:

50% of the cases: **Q1** (embarrassing)
50% of the cases: Q2 (regular, answer is always YES)

- Record responses $X_1, \ldots, X_{100}$, $X_i = 1$ (YES) or 0 (NO)

REMARKS ON RANDOMIZED RESPONSE

- Ask people two questions:

50% of the cases: **Q1** (embarrassing)
50% of the cases: Q2 (regular, answer is always YES)

- Record responses $X_1, \ldots, X_{100}$, $X_i = 1$ (YES) or 0 (NO)

$X_i \sim \text{Bernoulli}(q)$

$q = P(X_i = 1) = P(\text{YES}) =$

REMARKS ON RANDOMIZED RESPONSE

- Ask people two questions:

50% of the cases: **Q1** (embarrassing)
50% of the cases: Q2 (regular, answer is always YES)

- Record responses $X_1, \ldots, X_{100}$, $X_i = 1$ (YES) or 0 (NO)

$X_i \sim \text{Bernoulli}(q)$

$q = P(X_i = 1) = P(YES) = P(YES|Q1)*P(Q1) + P(YES|Q2)*P(Q2) =$

$=$

REMARKS ON RANDOMIZED RESPONSE

- Ask people two questions:

50% of the cases: **Q1** (embarrassing)
50% of the cases: Q2 (regular, answer is always YES)

- Record responses $X_1, \ldots, X_{100}$, $X_i = 1$ (YES) or 0 (NO)

$X_i \sim \text{Bernoulli}(q)$

$q = P(X_i = 1) = P(YES) = P(YES|Q1)*P(Q1) + P(YES|Q2)*P(Q2) =$
$= \mathbf{P(YES|Q1)}*0.5 + 0.5*1$

REMARKS ON RANDOMIZED RESPONSE

- Ask people two questions:

50% of the cases: **Q1** (embarrassing)
50% of the cases: Q2 (regular, answer is always YES)

- Record responses $X_1, \ldots, X_{100}$, $X_i = 1$ (YES) or 0 (NO)

$X_i \sim \text{Bernoulli}(q)$

$q = P(X_i = 1) = P(YES) = P(YES|Q1)*P(Q1) + P(YES|Q2)*P(Q2) =$
$= \mathbf{P(YES|Q1)}*0.5 + 0.5*1 = \mathbf{p}*0.5 + 0.5$

REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim$ Bernoulli(q)
- q = 0.5*$\mathbf{p}$ + 0.5

- MLE: $\hat{q}$ = {We got 60 YES out of 100} =

REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim$ Bernoulli(q)
- q = 0.5***p** + 0.5

- MLE: $\hat{q}$ = {We got 60 YES out of 100} = 0.6 $\Rightarrow \hat{p}$ =

REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim$ Bernoulli(q)
- $q = 0.5 * \mathbf{p} + 0.5$

- MLE: $\hat{q}$ = {We got 60 YES out of 100} = 0.6 $\Rightarrow \hat{p} = 2\hat{q} - 1 = 0.2$

REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim \text{Bernoulli}(q)$
- $q = 0.5^*\mathbf{p} + 0.5$

- MLE: $\hat{q}$ = {We got 60 YES out of 100} = 0.6 $\Rightarrow \hat{p} = 2\hat{q} - 1 = 0.2$

- Imagine that we got 40 YES out of 100. Then:

$\hat{q} =$ $\Rightarrow \hat{p} =$

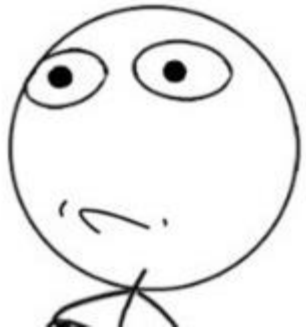
REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim \text{Bernoulli}(q)$
- $q = 0.5*\mathbf{p} + 0.5$

- MLE: $\hat{q}$ = {We got 60 YES out of 100} = 0.6 $\Rightarrow \hat{p} = 2\hat{q} - 1 = 0.2$

- Imagine that we got 40 YES out of 100. Then:

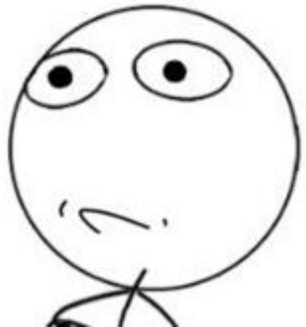
$$\hat{q} = 0.4 \qquad \Rightarrow \hat{p} = 2\hat{q} - 1 = -0.2$$



REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim \text{Bernoulli}(q)$
- $q = 0.5*\mathbf{p} + 0.5$
- MLE: $\hat{q}$ = {We got 60 YES out of 100} = 0.6 $\Rightarrow \hat{p} = 2\hat{q} - 1 = 0.2$
- Imagine that we got 40 YES out of 100. Then:

$$\cancel{\hat{q} = 0.4 \qquad \Rightarrow \hat{p} = 2\hat{q} - 1 = -0.2}$$



REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim$ Bernoulli(q)
- q = 0.5***p** + 0.5

REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim$ Bernoulli(q)
- q = 0.5***p** + 0.5

Constraints: $0 \le \hat{q} \le 1$,

$0 \le \hat{p} \le 1$

REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim \text{Bernoulli}(q)$
- $q = 0.5*\mathbf{p} + 0.5$

Constraints: $0 \leq \hat{q} \leq 1,$

$0 \leq \hat{p} \leq 1$ $\boxed{\Rightarrow 0.5 \leq \hat{q} \leq 1}$

REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim$ Bernoulli(q)
- q = 0.5***p** + 0.5

Constraints: $0 \leq \hat{q} \leq 1$,
$0 \leq \hat{p} \leq 1$

$$\boxed{\Rightarrow 0.5 \leq \hat{q} \leq 1}$$

- When maximizing a function with constraints: check the borders.

REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim$ Bernoulli(q)
- q = 0.5***p** + 0.5

Constraints: $0 \leq \hat{q} \leq 1,$ $\boxed{\Rightarrow 0.5 \leq \hat{q} \leq 1}$

$0 \leq \hat{p} \leq 1$

- When maximizing a function with constraints: check the borders.

- 40 out of 100 YES. MLE:

0.4 gives the highest L(q), but out of reach

L(0.5) L(1)

REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim$ Bernoulli(q)
- q = 0.5***p** + 0.5

Constraints: $0 \leq \hat{q} \leq 1,$ $\boxed{\Rightarrow 0.5 \leq \hat{q} \leq 1}$

$0 \leq \hat{p} \leq 1$

- When maximizing a function with constraints: check the borders.

- 40 out of 100 YES. MLE:

  0.4 gives the highest L(q), but out of reach

  L(0.5) > L(1) = 0

REMARKS ON RANDOMIZED RESPONSE

- Responses $X_1, \ldots, X_{100}$, $X_i \sim \text{Bernoulli}(q)$
- $q = 0.5*\mathbf{p} + 0.5$

Constraints: $0 \le \hat{q} \le 1$,

$0 \le \hat{p} \le 1$ $\boxed{\Rightarrow 0.5 \le \hat{q} \le 1}$

- When maximizing a function with constraints: check the borders.

- 40 out of 100 YES. MLE:

  0.4 gives the highest L(q), but out of reach

  $L(0.5) > L(1) = 0$ $\Rightarrow \boldsymbol{\hat{q} = 0.5}, \quad \boldsymbol{\hat{p} = 0}$

OVERVIEW
OF SOME DISTRIBUTIONS
WE WORKED WITH

DISTRIBUTIONS WE'VE SEEN SO FAR

- **Bernoulli**

X ~ Bernoulli(p) $\mathrm{P}(X = 1) = p, \quad P(X = 0) = 1 - p$

Chance of success in a single trial with two outcomes

DISTRIBUTIONS WE'VE SEEN SO FAR

- **Bernoulli**

X ~ Bernoulli(p) $\mathrm{P}(X = 1) = p, \quad P(X = 0) = 1 - p$

E(X) = p Chance of success in a single trial with two outcomes

DISTRIBUTIONS WE'VE SEEN SO FAR

- **Bernoulli**

X ~ Bernoulli(p) $\mathrm{P}(X = 1) = p, \quad P(X = 0) = 1 - p$

E(X) = p Chance of success in a single trial with two outcomes

- **Binomial**

X ~ Bi(n, p), $\mathrm{P}(X = k) = C_n^k p^k (1-p)^{n-k}, \;\; 0 \le k \le 1$

DISTRIBUTIONS WE'VE SEEN SO FAR

- **Bernoulli**

X ~ Bernoulli(p) $\mathrm{P}(X = 1) = p, \quad P(X = 0) = 1 - p$

E(X) = p Chance of success in a single trial with two outcomes

- **Binomial**

X ~ Bi(n, p), $\mathrm{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \;\; 0 \leq k \leq 1$

\# of successes in a series of n Bernoulli trials

DISTRIBUTIONS WE'VE SEEN SO FAR

- **Bernoulli**

X ~ Bernoulli(p) $\mathrm{P}(X = 1) = p, \quad P(X = 0) = 1 - p$

E(X) = p Chance of success in a single trial with two outcomes

- **Binomial**

X ~ Bi(n, p), $\mathrm{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \;\; 0 \leq k \leq 1$

E(X) = np # of successes in a series of n Bernoulli trials

DISTRIBUTIONS WE'VE SEEN SO FAR

- **Bernoulli**

X ~ Bernoulli(p) $\mathrm{P}(X = 1) = p, \quad P(X = 0) = 1 - p$

E(X) = p Chance of success in a single trial with two outcomes

- **Binomial**

X ~ Bi(n, p), $\mathrm{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \;\; 0 \leq k \leq 1$

E(X) = np # of successes in a series of n Bernoulli trials

- **Poisson**

X ~ Po(λ), $\mathrm{P}(X = k) = \frac{\mathrm{e}^{-\lambda} \cdot \lambda^{\mathrm{k}}}{\mathrm{k}!}, \;\; k \geq 0$

DISTRIBUTIONS WE'VE SEEN SO FAR

- **Bernoulli**

X ~ Bernoulli(p) $\mathrm{P}(X=1)=p, \quad P(X=0)=1-p$

E(X) = p Chance of success in a single trial with two outcomes

- **Binomial**

X ~ Bi(n, p), $\mathrm{P}(X=k)=C_n^k p^k(1-p)^{n-k}, \;\; 0 \leq k \leq 1$

E(X) = np # of successes in a series of n Bernoulli trials

- **Poisson**

X ~ Po(λ), $\mathrm{P}(X=k)=\frac{\mathrm{e}^{-\lambda} \cdot \lambda^{\mathrm{k}}}{\mathrm{k}!}, \;\; k \geq 0$

events that occur within a fixed amount of time

DISTRIBUTIONS WE'VE SEEN SO FAR

- **Bernoulli**

X ~ Bernoulli(p) $\mathrm{P}(X = 1) = p, \quad P(X = 0) = 1 - p$

E(X) = p Chance of success in a single trial with two outcomes

- **Binomial**

X ~ Bi(n, p), $\mathrm{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \;\; 0 \leq k \leq 1$

E(X) = np # of successes in a series of n Bernoulli trials

- **Poisson**

X ~ Po(λ), $\mathrm{P}(X = k) = \frac{\mathrm{e}^{-\lambda} \cdot \lambda^{\mathrm{k}}}{\mathrm{k}!}, \;\; k \geq 0$

E(X) = λ # events that occur within a fixed amount of time

MLE ONCE AGAIN

+ one more discrete distribution

WAITING FOR A METRO

- You are waiting for a train in the metro.
- Trains go every t minutes (no delays).

WAITING FOR A METRO

- You are waiting for a train in the metro.
- Trains go every t minutes (no delays).
- How long may you have to wait?

WAITING FOR A METRO

- You are waiting for a train in the metro.
- Trains go every *t* minutes (no delays).
- How long may you have to wait?
  - Anything between 0 and t minutes.

WAITING FOR A METRO

- You are waiting for a train in the metro.
- Trains go every *t* minutes (no delays).
- How long may you have to wait?
  - Anything between 0 and t minutes.
- You ask N of your friends how long have they waited:

$$T_1, T_2, \ldots, T_N$$

WAITING FOR A METRO

- You are waiting for a train in the metro.
- Trains go every *t* minutes (no delays).
- How long may you have to wait?
  - Anything between 0 and t minutes.
- You ask N of your friends how long have they waited:

$$T_1, T_2, \ldots, T_N$$

- How often do the trains go?
  And how much will *you* need to wait?

WAITING FOR A METRO

- Let's start with estimating the value of parameter t (train come every t minutes).

WAITING FOR A METRO

- Let's start with estimating the value of parameter t (train come every t minutes).

- Assume a very simple model: X – one's waiting time.

WAITING FOR A METRO

- Let's start with estimating the value of parameter *t* (train come every *t* minutes).

- Assume a very simple model: X – one's waiting time.

X	0	1	…	t
P(X)				

WAITING FOR A METRO

- Let's start with estimating the value of parameter *t* (train come every *t* minutes).

- Assume a very simple model: X – one's waiting time.

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

WAITING FOR A METRO

- Let's start with estimating the value of parameter *t* (train come every *t* minutes).

- Assume a very simple model: X – one's waiting time.

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

WAITING FOR A METRO

- Let's start with estimating the value of parameter *t* (train come every *t* minutes).

- Assume a very simple model: X – one's waiting time.

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- You know how long *N* of your friends waited: $T_1, T_2, \ldots, T_N$

  How to estimate the parameter t?

WAITING FOR A METRO

- Let's start with estimating the value of parameter *t* (train come every *t* minutes).

- Assume a very simple model: X – one's waiting time.

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- You know how long *N* of your friends waited: $T_1, T_2, \ldots, T_N$

  How to estimate the parameter t? -> **Maximum likelihood** ☺

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… *and 0 otherwise*

- Your friends waited for $T_1, T_2, \ldots, T_N$

maximize $L(t) =$ w.r.t. t

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- Your friends waited for $T_1, T_2, \ldots, T_N$

maximize $L(t) = \prod_{i=1}^{N} \frac{1}{t+1}$ w.r.t. t

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- Your friends waited for $T_1, T_2, \ldots, T_N$

maximize $$L(t) = \prod_{i=1}^{N} \frac{1}{t+1} \cdot \boldsymbol{I}(\boldsymbol{T_i} \leq \boldsymbol{t})$$ w.r.t. t

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- Your friends waited for $T_1, T_2, \ldots, T_N$

maximize $\quad L(t) = \prod_{i=1}^{N} \frac{1}{t+1} \cdot \boldsymbol{I(T_i \leq t)} \quad$ w.r.t. t

$$I(T_i \leq t) \neq 0 \Rightarrow t \geq \max(T_1, \ldots, T_N) = T$$

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- Your friends waited for $T_1, T_2, \ldots, T_N$

maximize $\quad L(t) = \prod_{i=1}^{N} \frac{1}{t+1} \cdot \boldsymbol{I(T_i \leq t)} \quad$ w.r.t. t

$$I(T_i \leq t) \neq 0 \Rightarrow t \geq \max(T_1, \ldots, T_N) = T$$

$$\log L(t) = \sum_{i=1}^{N} \log \frac{1}{t+1} = -N \log(t+1)$$

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- Your friends waited for $T_1, T_2, \ldots, T_N$

maximize $\quad L(t) = \prod_{i=1}^{N} \frac{1}{t+1} \cdot \boldsymbol{I(T_i \leq t)}$ $\quad$ w.r.t. t

$$I(T_i \leq t) \neq 0 \Rightarrow t \geq \max(T_1, \ldots, T_N) = T$$

$$\log L(t) = \sum_{i=1}^{N} \log \frac{1}{t+1} = -N \log(t+1)$$

decreasing function, need the smallest t possible $\Rightarrow \boldsymbol{\hat{t} = T = max(T_1, \ldots, T_N)}$

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- Your friends waited for $T_1, T_2, \ldots, T_N$
- **MLE:** $\hat{t} = T = max(T_1, \ldots, T_N)$

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- Your friends waited for $T_1, T_2, \ldots, T_N$
- **MLE:** $\hat{t} = T = max(T_1, \ldots, T_N)$
- Example: your friends waited for *0, 2, 4, 5, 10, 0, 2, 1, 3, 1* minutes.
- MLE:

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- Your friends waited for $T_1, T_2, \ldots, T_N$
- **MLE:** $\hat{\boldsymbol{t}} = \boldsymbol{T} = \boldsymbol{max}(\boldsymbol{T_1}, \ldots, \boldsymbol{T_N})$

- Example: your friends waited for *0, 2, 4, 5, 10, 0, 2, 1, 3, 1* minutes.
- MLE: trains every 10 minutes

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- Your friends waited for $T_1, T_2, \ldots, T_N$
- **MLE:** $\boldsymbol{\hat{t} = T = max(T_1, \ldots, T_N)}$

- Example: your friends waited for *0, 2, 4, 5, 10, 0, 2, 1, 3, 1* minutes.
- MLE: trains every 10 minutes
- Your friends waited on average 2.8 minutes

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… *and 0 otherwise*

- Your friends waited for $T_1, T_2, \ldots, T_N$
- **MLE:** $\hat{\boldsymbol{t}} = \boldsymbol{T} = \boldsymbol{max}(\boldsymbol{T_1}, \ldots, \boldsymbol{T_N})$
- How long will you need to wait *on average*?

$$E(X) =$$

- Example: your friends waited for *0, 2, 4, 5, 10, 0, 2, 1, 3, 1* minutes.
- MLE: trains every 10 minutes
- Your friends waited on average 2.8 minutes

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- Your friends waited for $T_1, T_2, \ldots, T_N$
- **MLE:** $\boldsymbol{\hat{t} = T = max(T_1, \ldots, T_N)}$
- How long will you need to wait *on average*?

$$E(X) = \sum_{k=0}^{\hat{t}} k \cdot \frac{1}{\hat{t}+1} =$$

- Example: your friends waited for *0, 2, 4, 5, 10, 0, 2, 1, 3, 1* minutes.
- MLE: trains every 10 minutes
- Your friends waited on average 2.8 minutes

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- Your friends waited for $T_1, T_2, \ldots, T_N$
- **MLE:** $\hat{\boldsymbol{t}} = \boldsymbol{T} = \boldsymbol{max}(\boldsymbol{T_1}, \ldots, \boldsymbol{T_N})$
- How long will you need to wait *on average*?

$$E(X) = \sum_{k=0}^{\hat{t}} k \cdot \frac{1}{\hat{t}+1} = \frac{\hat{t}}{2}$$

- Example: your friends waited for *0, 2, 4, 5, 10, 0, 2, 1, 3, 1* minutes.
- MLE: trains every 10 minutes
- Your friends waited on average 2.8 minutes

WAITING FOR A METRO

X	0	1	…	t
P(X)	1/(t+1)	1/(t+1)	…	1/(t+1)

… and 0 otherwise

- Your friends waited for $T_1, T_2, \ldots, T_N$
- **MLE:** $\boldsymbol{\hat{t} = T = max(T_1, \ldots, T_N)}$
- How long will you need to wait *on average*?

$$E(X) = \sum_{k=0}^{\hat{t}} k \cdot \frac{1}{\hat{t}+1} = \frac{\hat{t}}{2}$$

- Example: your friends waited for *0, 2, 4, 5, 10, 0, 2, 1, 3, 1* minutes.
- MLE: trains every 10 minutes
- Your friends waited on average 2.8 minutes
- You'll have to wait 5 minutes

DISCRETE UNIFORM DISTRIBUTION

- Takes values from 1 to n with equal probabilities:

- **Typical example:** *rolling a fair die*

DISCRETE UNIFORM DISTRIBUTION

- Takes values from 1 to n with equal probabilities:

$$P(X = k) = \begin{cases} \frac{1}{n}, & 1 \le k \le n \\ 0, & otherwise \end{cases}$$

$$E(X) =$$

- **Typical example:** *rolling a fair die*

DISCRETE UNIFORM DISTRIBUTION

- Takes values from 1 to n with equal probabilities:

$$P(X = k) = \begin{cases} \frac{1}{n}, & 1 \leq k \leq n \\ 0, & otherwise \end{cases}$$

$$E(X) = \sum_{k=1}^{n} \frac{k}{n} = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}$$

- **Typical example:** *rolling a fair die*

DISCRETE UNIFORM DISTRIBUTION

- Takes values from 1 to n with equal probabilities:

$$P(X=k)=\begin{cases}\frac{1}{n}, & 1\le k\le n\\ 0, & otherwise\end{cases}$$

$$E(X)=\sum_{k=1}^{n}\frac{k}{n}=\frac{1}{n}\cdot\frac{n(n+1)}{2}=\frac{n+1}{2}$$

- **Typical example:** *rolling a fair die*

- And if we roll two dice and consider a sum of the values we get?

LOOKING
FOR A BETTER MODEL

WE NEED A BETTER MODEL

- A model used in the previous example was too simple: waiting time can be not just *0, 1, 2, … t* minutes.
- We need a distribution that takes *more* values.

WE NEED A BETTER MODEL

- A model used in the previous example was too simple: waiting time can be not just *0, 1, 2, … t* minutes.

- We need a distribution that takes *more* values.

- In fact, we need out random variable to take *infinite* number of values…

DIFFERENT KINDS OF ‘INFINITE’

FINITE SET OF VALUES

- X ~ Bernoulli(p)

- X ~ Bi(n, p)

- X ~ Uniform discrete (k)

DIFFERENT KINDS OF 'INFINITE'

FINITE SET OF VALUES

- X ~ Bernoulli(p)
  - 2 values
- X ~ Bi(n, p)
  - n+1 values
- X ~ Uniform discrete (k)
  - k (or k+1 as in our example) values

DIFFERENT KINDS OF 'INFINITE'

FINITE SET OF VALUES

- X ~ Bernoulli(p)
  - 2 values
- X ~ Bi(n, p)
  - n+1 values
- X ~ Uniform discrete (k)
  - k (or k+1 as in our example) values

INFINITE SET OF VALUES

- X ~ Po(λ)
  - 0, 1, 2, 3, …
  - countable set of values

DIFFERENT KINDS OF 'INFINITE'

FINITE SET OF VALUES

- X ~ Bernoulli(p)
  - 2 values
- X ~ Bi(n, p)
  - n+1 values
- X ~ Uniform discrete (k)
  - k (or k+1 as in our example) values

INFINITE SET OF VALUES

- X ~ Po(λ)
  - 0, 1, 2, 3, …
  - countable set of values
- ?
  - uncountably many values

DIFFERENT KINDS OF 'INFINITE'

FINITE SET OF VALUES

- X ~ Bernoulli(p)
  - 2 values
- X ~ Bi(n, p)
  - n+1 values
- X ~ Uniform discrete (k)
  - k (or k+1 as in our example) values

INFINITE SET OF VALUES

- X ~ Po(λ)
  - 0, 1, 2, 3, …
  - countable set of values
- **Continuous distributions**
  - uncountably many values

CONTINUOUS UNIFORM DISTRIBUTION

UNIFORM DISTRIBUTION

- We want to model a random variable X that takes *any* value between *a* and *b* with equal probability.

UNIFORM DISTRIBUTION

- We want to model a random variable X that takes *any* value between *a* and *b* with equal probability.

- For example, let *a* = *0* and *b* = *10*:

UNIFORM DISTRIBUTION

- We want to model a random variable X that takes *any* value between *a* and *b* with equal probability.

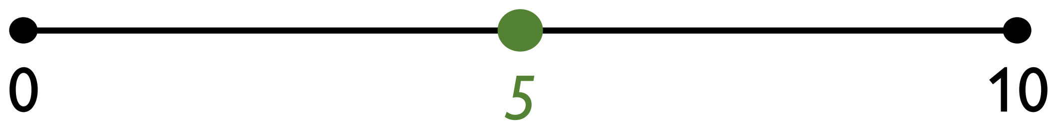
- For example, let $a = 0$ and $b = 10$:

- What's the expected value of such a variable?

UNIFORM DISTRIBUTION

- We want to model a random variable X that takes *any* value between *a* and *b* with equal probability.

- For example, let *a* = *0* and *b* = *10*:

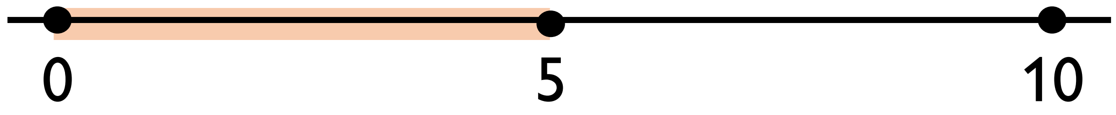


- What's the expected value of such a variable? **E(X) = (10 - 0)/2 = 5**

UNIFORM DISTRIBUTION

- What's the probability $P(X \leq 5)$?

- $P(X \leq 5) =$

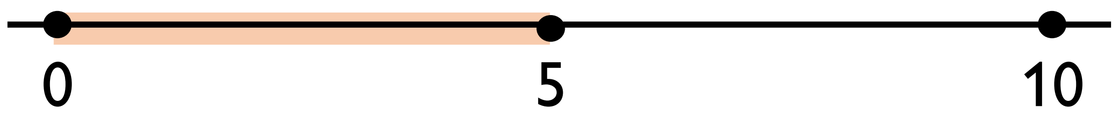


UNIFORM DISTRIBUTION

- What's the probability $P(X \leq 5)$?

- $P(X \leq 5) = (5 - 0) / (10 - 0) =$

  $= 0.5$



UNIFORM DISTRIBUTION

- What's the probability $P(X \leq 3)$?

- $P(X \leq 3) =$

UNIFORM DISTRIBUTION

- What's the probability $P(X \leq 3)$?

- $P(X \leq 3) = (3 - 0) / (10 - 0) =$

  $= 0.3$

UNIFORM DISTRIBUTION

- What's the probability $P(X \leq -1)$?

- $P(X \leq -1) =$

UNIFORM DISTRIBUTION

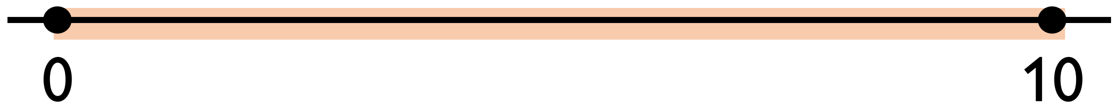
- What's the probability $P(X \leq -1)$?

- $P(X \leq -1) = 0$

UNIFORM DISTRIBUTION

- What's the probability $P(X \leq 15)$?

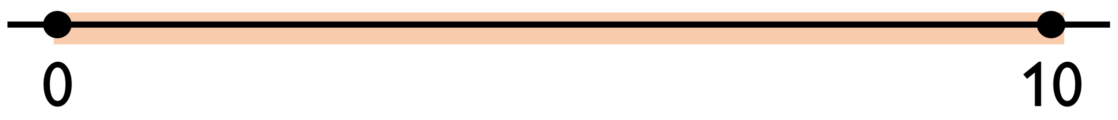
- $P(X \leq 15) =$



UNIFORM DISTRIBUTION

- What's the probability $P(X \leq 15)$?

- $P(X \leq 15) = 1$



UNIFORM DISTRIBUTION

- CUMULATIVE DISTRIBUTION FUNCTION (CDF):

$$\mathbf{F(x) = P(X \leq x)}$$

UNIFORM DISTRIBUTION

- CUMULATIVE DISTRIBUTION FUNCTION (CDF):

$$\mathbf{F(x) = P(X \leq x)}$$

- Uniform distribution:

$$\mathrm{F(x)} = \begin{cases} 0, & x < a \\ & \\ & \end{cases}$$

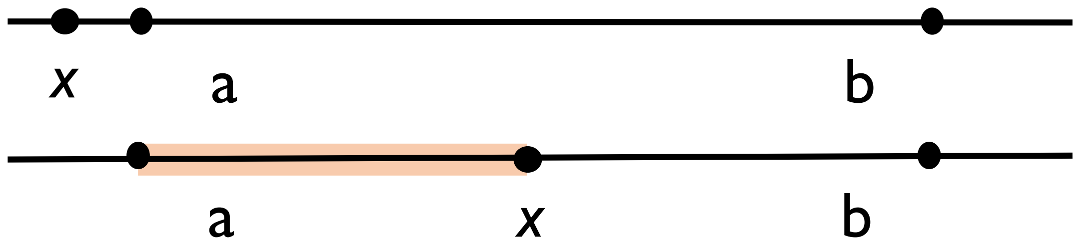
UNIFORM DISTRIBUTION

- CUMULATIVE DISTRIBUTION FUNCTION (CDF):

$$\mathbf{F(x) = P(X \leq x)}$$

- Uniform distribution:

$$F(x) = \begin{cases} 0, & x < a \\ & \\ & \end{cases}$$



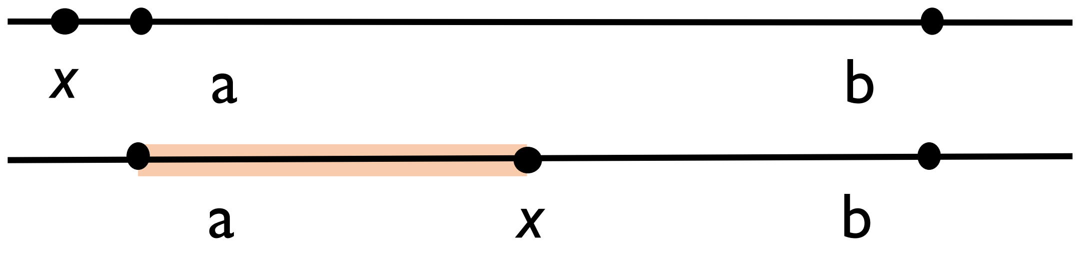
UNIFORM DISTRIBUTION

- CUMULATIVE DISTRIBUTION FUNCTION (CDF):

$$\mathbf{F(x) = P(X \leq x)}$$

- Uniform distribution:

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \end{cases}$$



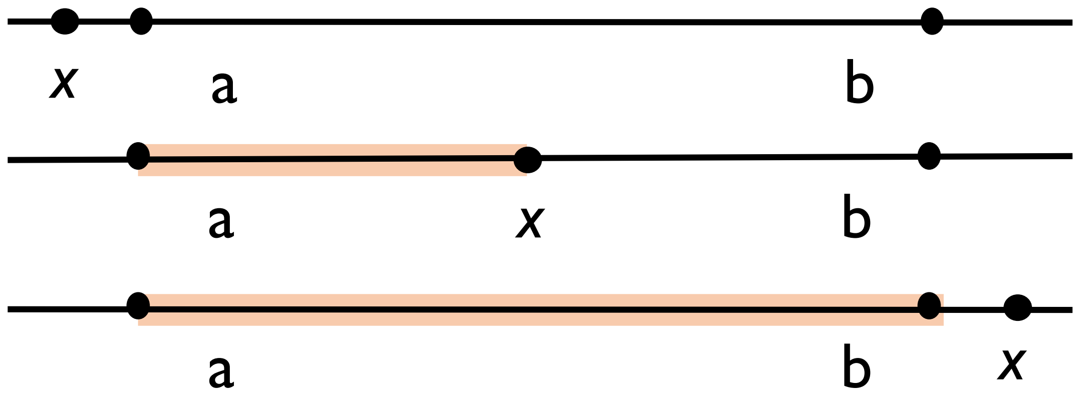
UNIFORM DISTRIBUTION

- CUMULATIVE DISTRIBUTION FUNCTION (CDF):

$$\mathbf{F(x) = P(X \leq x)}$$

- Uniform distribution:

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x - a}{b - a}, & a \leq x \leq b \end{cases}$$



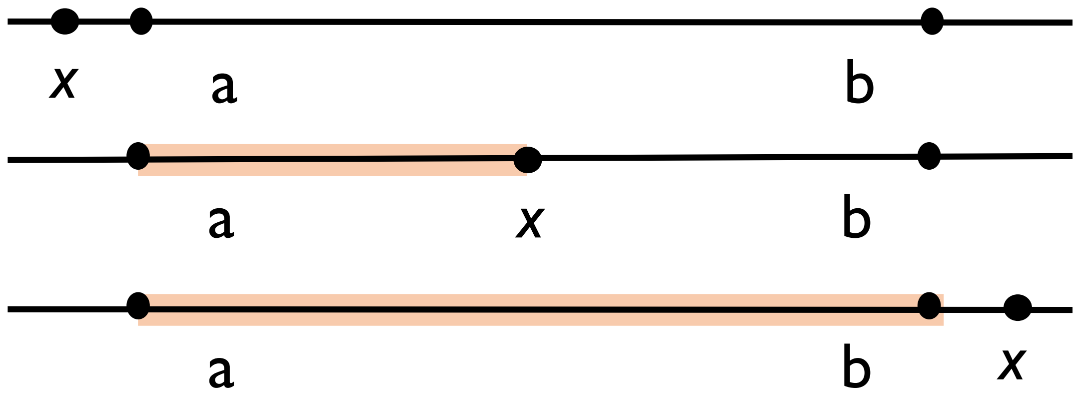
UNIFORM DISTRIBUTION

- CUMULATIVE DISTRIBUTION FUNCTION (CDF):

$$\mathbf{F(x) = P(X \leq x)}$$

- Uniform distribution:

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$



CDF

- Cumulative Distribution Function:

$$\mathbf{F(x) = P(X \leq x)}$$

- Which basic properties such a function has?

CDF

- Cumulative Distribution Function:

$$\mathbf{F(x) = P(X \leq x)}$$

- Which basic properties such a function has?
  - $0 \leq F(x) \leq 1$
  - $F(x)$ is non-decreasing

CDF

- Cumulative Distribution Function:

$$\mathbf{F(x) = P(X \leq x)}$$

- Which basic properties such a function has?
  - $0 \leq F(x) \leq 1$
  - $F(x)$ is non-decreasing

- CDF defines a continuous distribution!

UNIFORM DISTRIBUTION

- What's the probability $P(X > 7)$?
- $P(X > 7) =$

UNIFORM DISTRIBUTION

- What's the probability $P(X > 7)$?

- $P(X > 7) = 1 - P(X \leq 7) =$

UNIFORM DISTRIBUTION

- What's the probability $P(X > 7)$?

- $P(X > 7) = 1 - P(X \leq 7) =$
  $= 1 - F(7) =$

UNIFORM DISTRIBUTION

- What's the probability $P(X > 7)$?

- $P(X > 7) = 1 – P(X \leq 7) =$

  $= 1 – F(7) =$

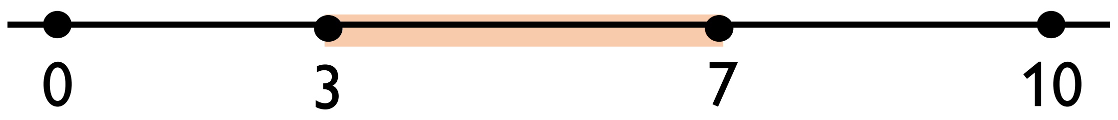
  $= 1 – (7 - 0)/(10 - 0) =$

  $= 1 – 0.7 = 0.3$

UNIFORM DISTRIBUTION

- What's the probability $P(3 < X \leq 7)$?

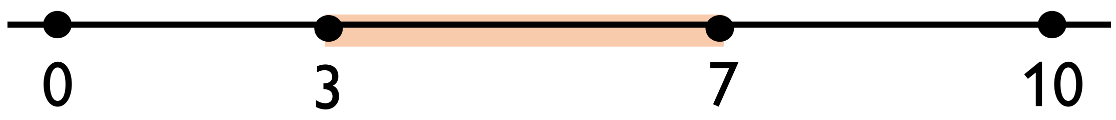
- $P(3 < X \leq 7) =$



UNIFORM DISTRIBUTION

- What's the probability $P(3 < X \leq 7)$?

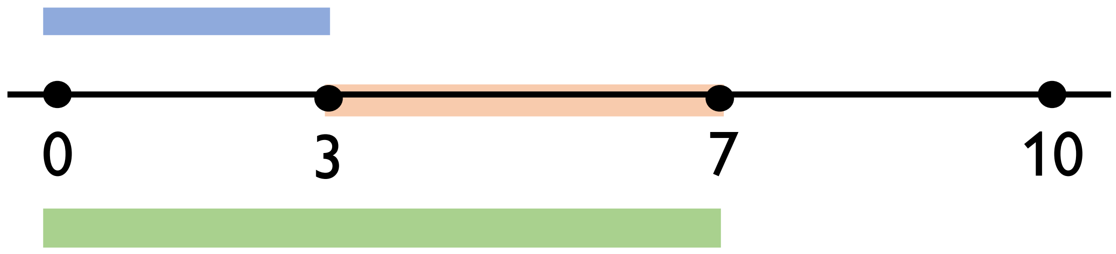
- $P(3 < X \leq 7) = (7 - 3)/(10 - 0) =$
  $= 0.4$



UNIFORM DISTRIBUTION

- What's the probability $P(3 < X \leq 7)$?

- $P(3 < X \leq 7) = (7 - 3)/(10 - 0) =$
  $= 0.4$



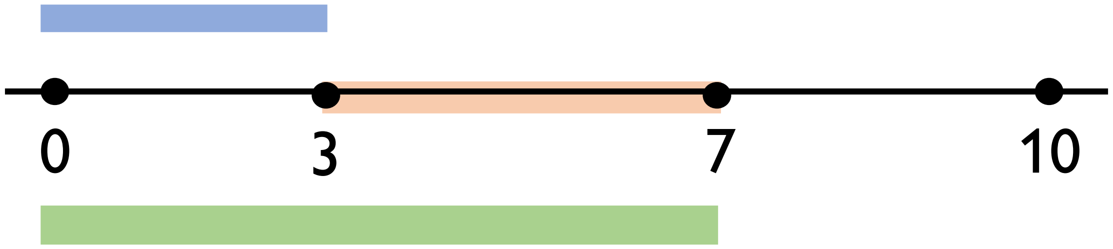
In terms of CDF:

$P(3 \leq X \leq 7) =$

UNIFORM DISTRIBUTION

- What's the probability $P(3 < X \leq 7)$?

- $P(3 < X \leq 7) = (7 - 3)/(10 - 0) =$
  $= 0.4$



In terms of CDF:

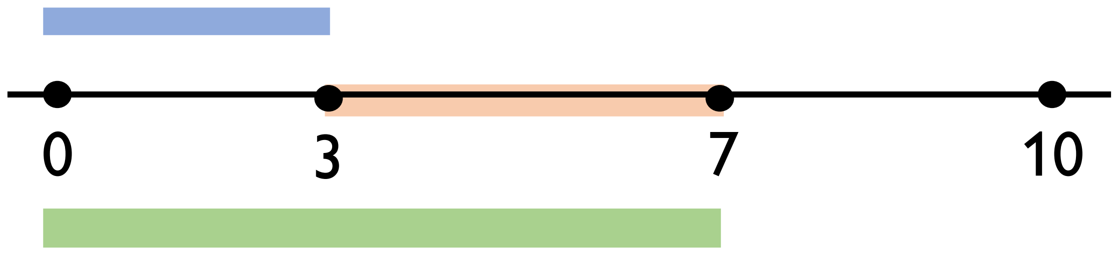
$P(3 \leq X \leq 7) =$

$= P(X \leq 7) - P(X \leq 3) =$

UNIFORM DISTRIBUTION

- What's the probability $P(3 < X \leq 7)$?

- $P(3 < X \leq 7) = (7 - 3)/(10 - 0) =$
  $= 0.4$



In terms of CDF:

$P(3 \leq X \leq 7) =$

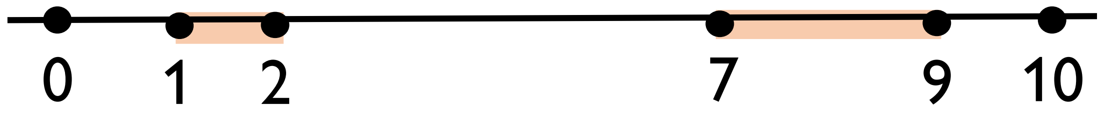
$= P(X \leq 7) - P(X \leq 3) =$

$= F(7) - F(3)$

UNIFORM DISTRIBUTION

- What's the probability $P(1 < X \leq 2 \text{ or } 7 < X \leq 9)$?

- $P(1 < X \leq 2 \text{ or } 7 < X \leq 9) =$



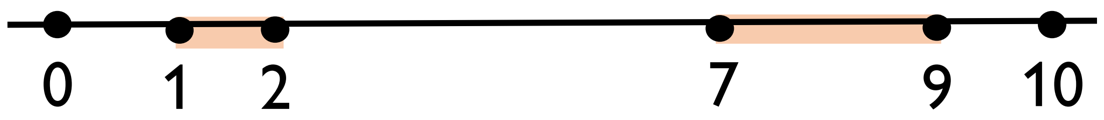
UNIFORM DISTRIBUTION

- What's the probability $P(1 < X \leq 2 \text{ or } 7 < X \leq 9)$?

- $P(1 < X \leq 2 \text{ or } 7 < X \leq 9) =$

  $= F(2) - F(1) +$

  $+ F(9) - F(7) =$



UNIFORM DISTRIBUTION

- What's the probability $P(1 < X \leq 2 \text{ or } 7 < X \leq 9)$?

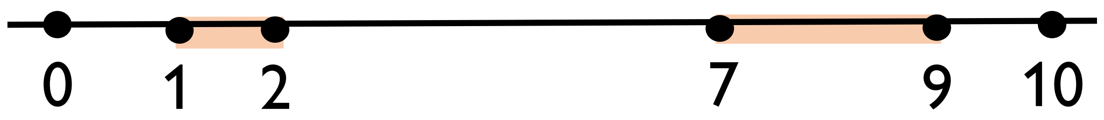
- $P(1 < X \leq 2 \text{ or } 7 < X \leq 9) =$

  $= F(2) - F(1) +$

  $+ F(9) - F(7) =$

  $= 0.2 - 0.1 + 0.9 - 0.7 =$

  $= 0.3$



CDF

We’ve figured out so far that

- $P(X \leq x) = F(x)$

CDF

We’ve figured out so far that

- $P(X \leq x) = F(x)$
- $P(X > x) =$

CDF

We’ve figured out so far that

- $P(X \leq x) = F(x)$
- $P(X > x) = 1 - F(x)$

CDF

We've figured out so far that

- $P(X \leq x) = F(x)$
- $P(X > x) = 1 - F(x)$
- $P(x1 < X \leq x2) =$

CDF

We've figured out so far that

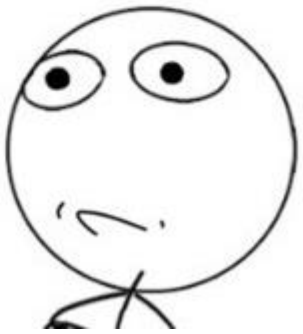
- $P(X \leq x) = F(x)$
- $P(X > x) = 1 - F(x)$
- $P(x1 < X \leq x2) = F(x2) - F(x1)$

CDF

We've figured out so far that

- $P(X \leq x) = F(x)$
- $P(X > x) = 1 - F(x)$
- $P(x1 < X \leq x2) = F(x2) - F(x1)$

But what's, for example, $P(X = 5)$?



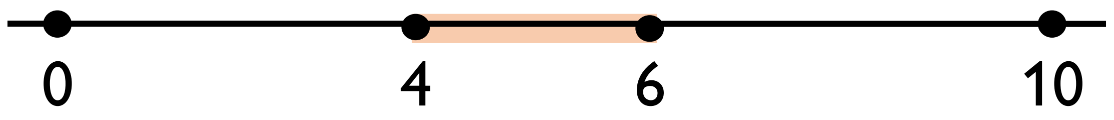
UNIFORM DISTRIBUTION

- $P(4 < X \leq 6) =$

UNIFORM DISTRIBUTION

- $P(4 < X \leq 6) =$

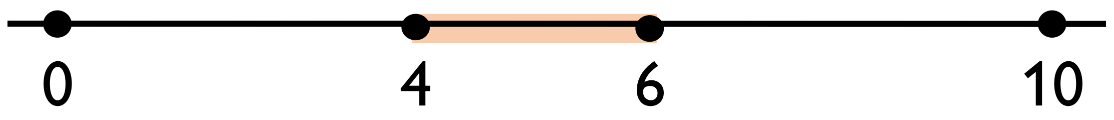
  $= (6 - 4)/(10 - 0) = 0.2$



UNIFORM DISTRIBUTION

- $P(4 < X \leq 6) =$

  $= (6 - 4)/(10 - 0) = 0.2$

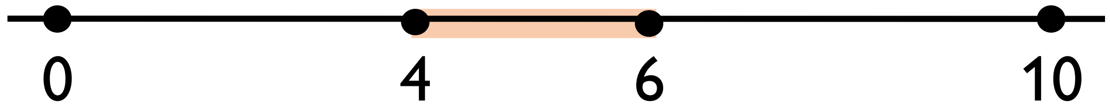


- $P(4.5 < X \leq 5.5) =$

UNIFORM DISTRIBUTION

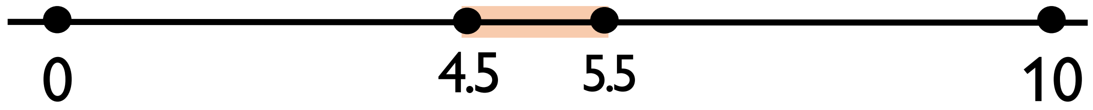
- $P(4 < X \leq 6) =$

  $= (6 – 4)/(10 - 0) = 0.2$



- $P(4.5 < X \leq 5.5) =$

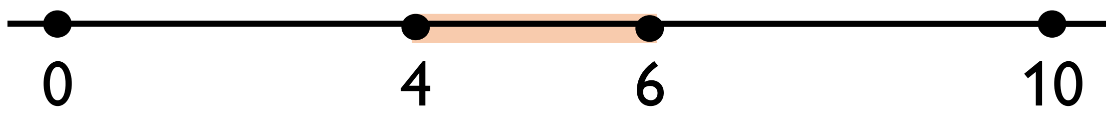
  $= (4.5 – 5.5)/(10 – 0) = 0.1$



UNIFORM DISTRIBUTION

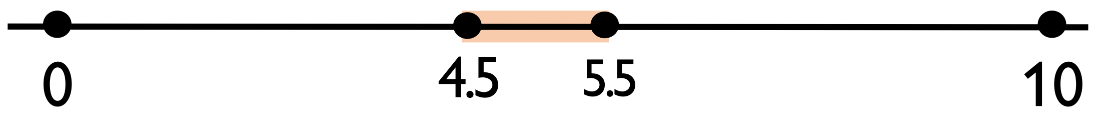
- $P(4 < X \leq 6) =$

  $= (6 – 4)/(10 - 0) = 0.2$



- $P(4.5 < X \leq 5.5) =$

  $= (4.5 – 5.5)/(10 – 0) = 0.1$

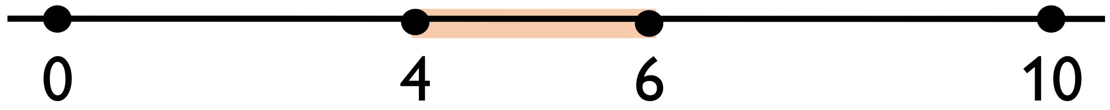


- $P(4.9 < X \leq 5.1) =$

UNIFORM DISTRIBUTION

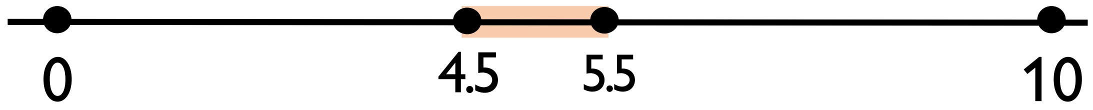
- $P(4 < X \leq 6) =$

  $= (6 – 4)/(10 - 0) = 0.2$



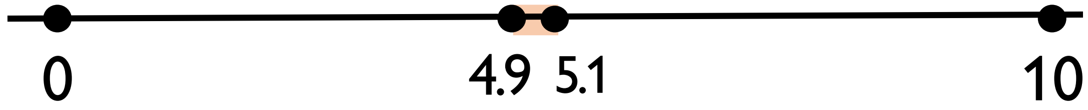
- $P(4.5 < X \leq 5.5) =$

  $= (4.5 – 5.5)/(10 – 0) = 0.1$



- $P(4.9 < X \leq 5.1) =$

  $= (4.9 – 5.1)/(10 – 0) = 0.02$



- $P(X = 5) = \ldots ?$

UNIFORM DISTRIBUTION

- We want to model a random variable X that takes *any* value between *0* and *10* with equal probability.

- But what's, for example, $P(X = 5)$?

UNIFORM DISTRIBUTION

- We want to model a random variable X that takes *any* value between *0* and *10* with equal probability.

- But what's, for example, $P(X = 5)$?

- Let's start with the discrete uniform distribution between 0 and 10: X takes n values between 0 and 10 with probability $p = 1/n$:

n	2	3					
1/p	*0.5*	*0.33*					

UNIFORM DISTRIBUTION

- We want to model a random variable X that takes *any* value between *0* and *10* with equal probability.

- But what's, for example, $P(X = 5)$?

- Let's start with the discrete uniform distribution between 0 and 10: X takes n values between 0 and 10 with probability $p = 1/n$:

n	2	3	10	1000	100000	…	
1/p	*0.5*	*0.33*	*0.1*	*0.001*	*0.00001*	…	

UNIFORM DISTRIBUTION

- We want to model a random variable X that takes *any* value between *0* and *10* with equal probability.

- But what's, for example, $P(X = 5)$?

- Let's start with the discrete uniform distribution between 0 and 10: X takes n values between 0 and 10 with probability $p = 1/n$:

n	2	3	10	1000	100000	…	∞
1/p	*0.5*	*0.33*	*0.1*	*0.001*	*0.00001*	…	*0*

Continuous random variables can take uncountably many values, but…

The probability that a continuous random variable is equal to a particular value is 0!

- So far, we've learned that:
  - Continuous random variables take uncountably many values.
  - If X is a continuous random variable, $P(X = x) = 0$.
  - Continuous distribution can be defined with a CDF.

CDFs FOR DISCRETE DISTRIBUTIONS

- Actually, CDF can be defined also for a *discrete* random variable.
- X – discrete, takes values x_1, x_2, … x_n

$$F(x) =$$

CDFs FOR DISCRETE DISTRIBUTIONS

- Actually, CDF can be defined also for a *discrete* random variable.
- X – discrete, takes values x_1, x_2, … x_n

$$F(x) = P(X \leq x) = \sum_{i=1}^{n} I(x_i \leq x) \cdot P(X = x_i)$$

CDFs FOR DISCRETE DISTRIBUTIONS

- Actually, CDF can be defined also for a *discrete* random variable.
- X – discrete, takes values x_1, x_2, … x_n

$$F(x) = P(X \le x) = \sum_{i=1}^{n} I(x_i \le x) \cdot P(X = x_i)$$

- What's CDF for Bernoulli distribution?

CDFs FOR DISCRETE DISTRIBUTIONS

- Actually, CDF can be defined also for a *discrete* random variable.
- X – discrete, takes values $x_1, x_2, \ldots x_n$

$$F(x) = P(X \leq x) = \sum_{i=1}^{n} I(x_i \leq x) \cdot P(X = x_i)$$

- What's CDF for Bernoulli distribution?

$$F(x) = P(X \leq x) =$$

CDFs FOR DISCRETE DISTRIBUTIONS

- Actually, CDF can be defined also for a *discrete* random variable.
- X – discrete, takes values x_1, x_2, … x_n

$$F(x) = P(X \leq x) = \sum_{i=1}^{n} I(x_i \leq x) \cdot P(X = x_i)$$

- What's CDF for Bernoulli distribution?

$$F(x) = P(X \leq x) = \begin{cases} & x < 0 \\ & 0 \leq x < 1 \\ & x \geq 1 \end{cases}$$

CDFs FOR DISCRETE DISTRIBUTIONS

- Actually, CDF can be defined also for a *discrete* random variable.
- X – discrete, takes values $x_1, x_2, \ldots x_n$

$$F(x) = P(X < x) = \sum_{i=1}^{n} I(x_i < x) \cdot P(X = x_i)$$

- What's CDF for Bernoulli distribution?

$$F(x) = P(X \le x) = \begin{cases} 0, & x < 0 \\ & 0 \le x < 1 \\ & x \ge 1 \end{cases}$$

CDFs FOR DISCRETE DISTRIBUTIONS

- Actually, CDF can be defined also for a *discrete* random variable.
- X – discrete, takes values x_1, x_2, … x_n

$$F(x) = P(X < x) = \sum_{i=1}^{n} I(x_i < x) \cdot P(X = x_i)$$

- What's CDF for Bernoulli distribution?

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ 1 - p, & 0 \leq x < 1 \\ & x \geq 1 \end{cases}$$

CDFs FOR DISCRETE DISTRIBUTIONS

- Actually, CDF can be defined also for a *discrete* random variable.
- X – discrete, takes values x_1, x_2, … x_n

$$F(x) = P(X < x) = \sum_{i=1}^{n} I(x_i < x) \cdot P(X = x_i)$$

- What's CDF for Bernoulli distribution?

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ 1 - p, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

EXPERIMENT WITH CDFs IN PYTHON

Google Classroom -> Probability mass functions and CDFs

TO SUM UP

- MLE for discrete uniform distribution
- Discrete vs continuous random variables
- CDF
  - for discrete
  - for continuous