



Geometric distributions

Back to Basics

Independent B_p coin flips

$$p(1) = p$$

$$p(0) = 1-p \stackrel{\text{def}}{=} q$$

Two Derived
Distributions

Binomial	$B_{p,n}$	n flips, # 1's
Geometric	G_p	# flips till first 1

Flips	X
10101	1
01011	2
00011	4

n	X_1, \dots, X_n	$p(n)$
1	$X_1 = 1$	p
2	$X_1 = 0 \quad X_2 = 1$	qp
3	$X_1 = X_2 = 0 \quad X_3 = 1$	q^2p
n	$X_1 = \dots = X_{n-1} = 0 \quad X_n = 1$	$q^{n-1}p$

Geometric Distribution

G_p

$$0 < p \leq 1$$

$$p(n) = q^{n-1}p \stackrel{\text{def}}{=} g_p(n) \quad n \geq 1$$

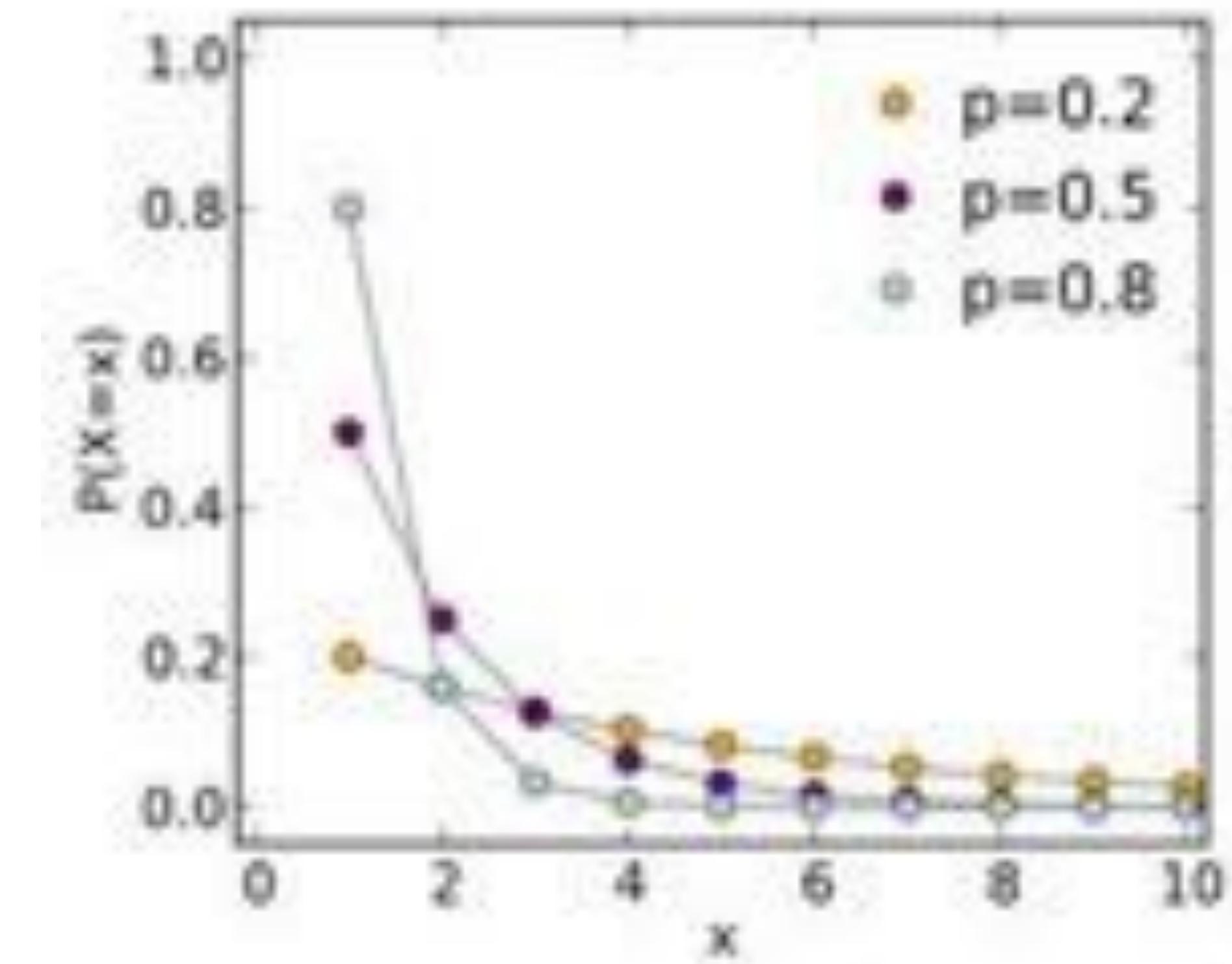
Notebook

Plot different p

Observations

$$p \neq 0$$

n can be arbitrarily high



Who's Geometric

30 years ago

Thief trying to find a matching key

Trials to hit a target

Attempts till success

Till failure

Nowadays



Σ WILL IT ADD?

$$P(n) = pq^{n-1} \quad n \geq 1 \quad q = 1 - p$$

$$(1 + q + q^2 + \dots)(1 - q) = 1 + \cancel{q} + \cancel{q^2} + \dots$$
$$\qquad\qquad\qquad - \cancel{q} - \cancel{q^2} - \dots$$

$$\sum_{i=0}^{\infty} q^i = \frac{1}{1 - q}$$

YES IT
ADDS!

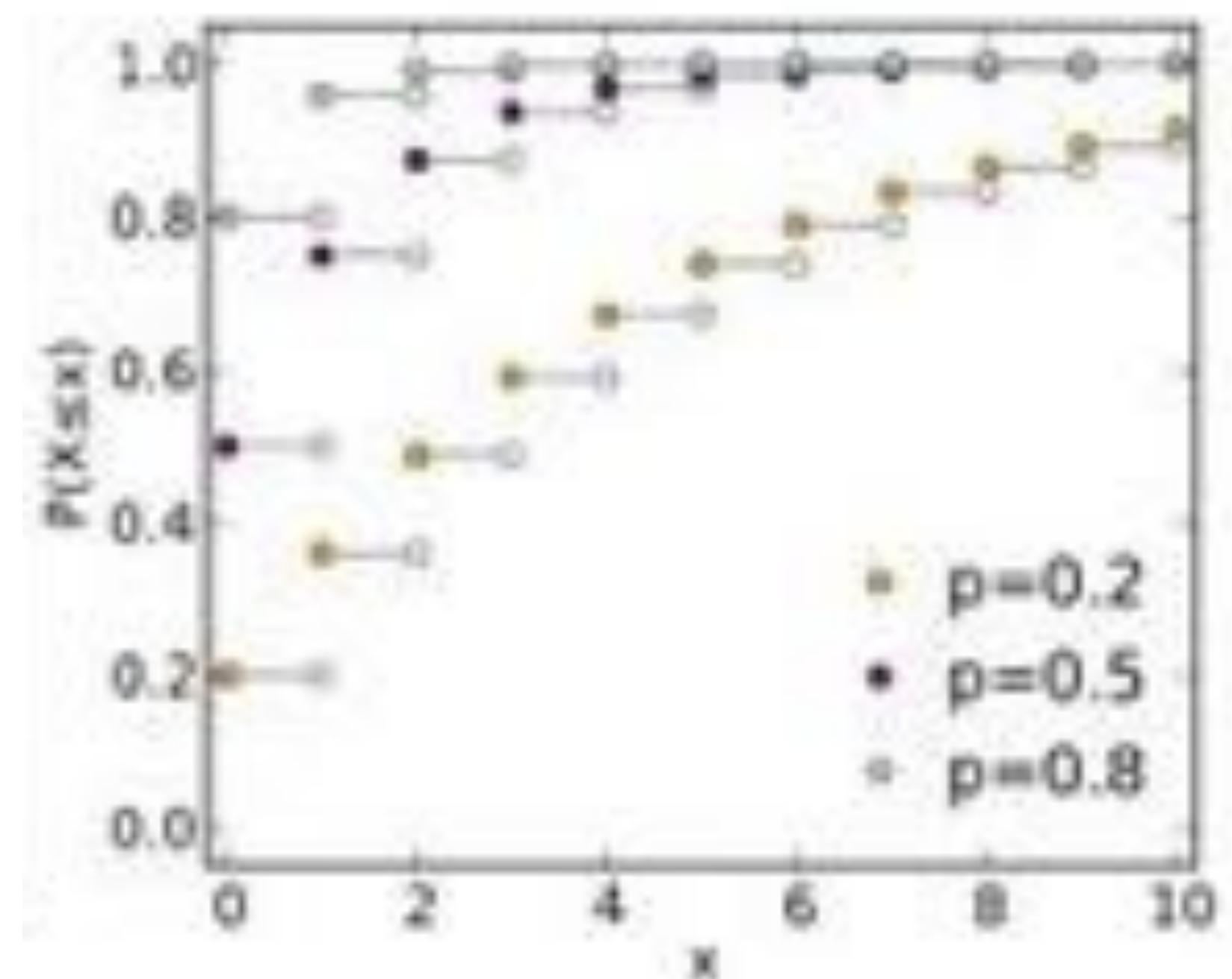
$$\sum_{n=1}^{\infty} p(1 - p)^{n-1} = p \sum_{i=0}^{\infty} (1 - p)^i = p \cdot \frac{1}{1 - (1 - p)} = \frac{p}{p} = 1$$

CDF

$n \in \mathbb{N}$

$$P(X > n) = P(X_1 = \dots = X_n = 0) = q^n$$

$$F(n) = P(X \leq n) = 1 - P(X > n) = 1 - q^n$$

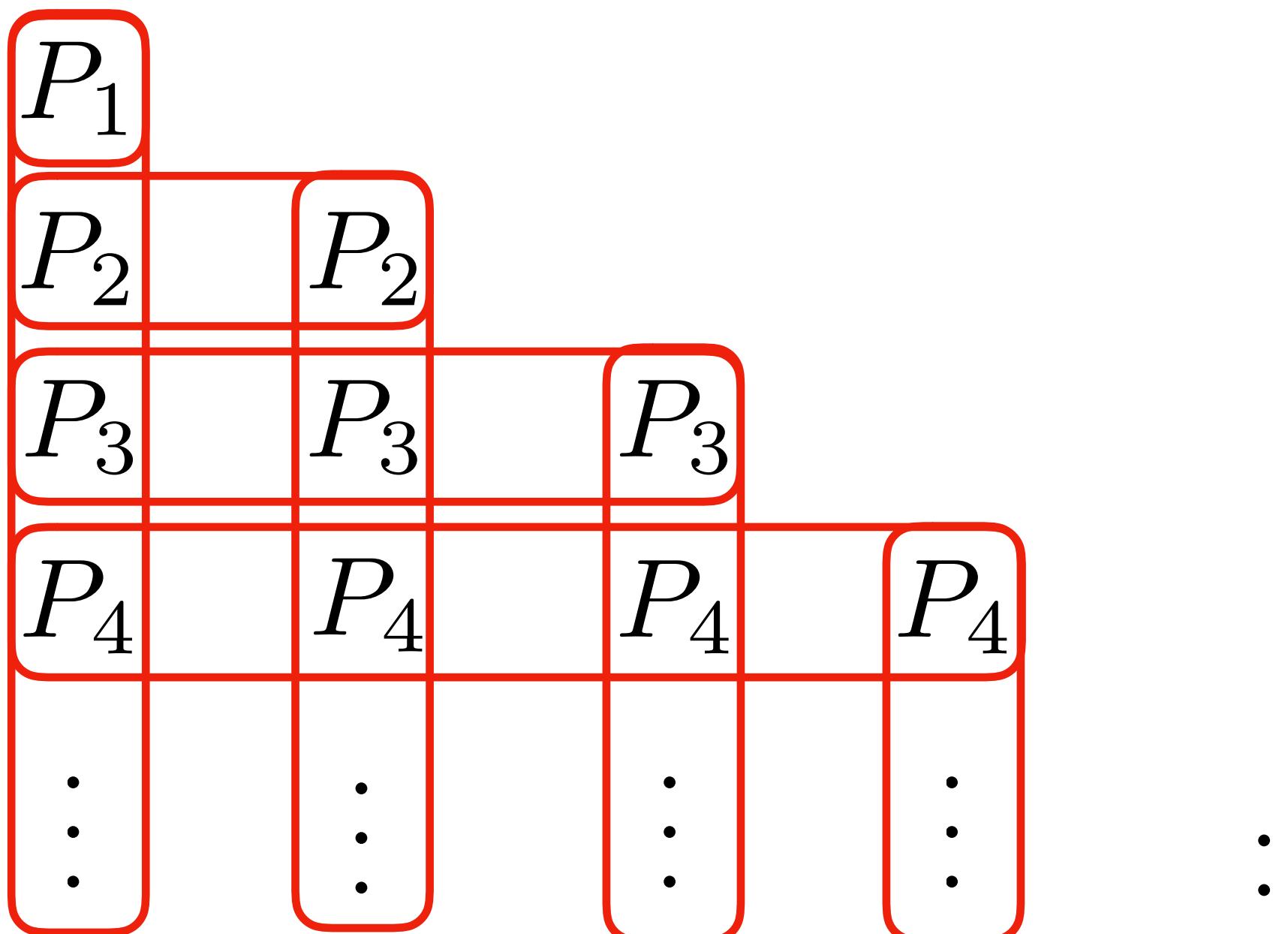


Expectation via “Right” CDF

$$x \in \mathbb{N} \quad P_k = P(X = k)$$

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} kP_k \\ &= P_1 + 2P_2 + 3P_3 + \dots \end{aligned}$$

$$= P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + \dots$$



Geometric distribution

$$EX = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{i=0}^{\infty} P(X > i) = \sum_{i=0}^{\infty} q^i = \frac{1}{1-q} = \frac{1}{p}$$

Variance

$$\begin{aligned} EX(X - 1) &= \sum_{n=1}^{\infty} n(n-1) \cdot P(X = n) = p \sum_{n=2}^{\infty} n(n-1)q^{n-1} \\ &= pq \sum_{n=2}^{\infty} \frac{d^2}{dq^2} q^n = pq \frac{d^2}{dq^2} \sum_{n=2}^{\infty} q^n \\ &= pq \frac{d^2}{dq^2} \left(\frac{1}{1-q} - 1 - q \right) \\ &= pq \frac{2}{(1-q)^3} = \frac{2q}{p^2} \end{aligned}$$

$$\begin{aligned} \left(\frac{1}{1-q} \right)' &= \frac{1}{(1-q)^2} \\ \left(\frac{1}{(1-q)^2} \right)' &= \frac{2}{(1-q)^3} \end{aligned}$$

$$EX^2 = EX(X - 1) + EX = \frac{2q}{p^2} + \frac{1}{p} = \frac{2q+p}{p^2} = \frac{1+q}{p^2}$$

$$V(X) = EX^2 - (EX)^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} \quad \sigma = \frac{\sqrt{q}}{p}$$

Fair Coin

$$X \sim G_{\frac{1}{2}}$$

$$P(X = k) = g_{0.5}(k) = \left(\frac{1}{2}\right)^{k-1} \cdot \frac{1}{2} = \frac{1}{2^k}$$

$$EX = \frac{1}{p} = 2$$

$$VX = \frac{q}{p^2} = 2$$

Memoryless

A distribution over $\mathbb{P} = \{1, 2, \dots\}$ is memoryless if for all $n \geq 0, m > 1$

$$P(X=n+m \mid X>n) = P(X=m)$$

$$P(X=12 \mid X>10) = P(X=2)$$

Geometric → Memoryless

$$P(X = n + m | X > n) = \frac{P(X=n+m, X>n)}{P(X>n)}$$

$$= \frac{P(X=n+m)}{P(X>n)}$$

$$= \frac{p \cdot q^{n+m-1}}{q^n}$$

$$= p \cdot q^{m-1}$$

$$= P(X = m)$$

All geometric distributions are memoryless

Memoryless → Geometric

Any discrete memoryless distribution over \mathbb{P} is geometric

Let $p \stackrel{\text{def}}{=} P(X = 1), \quad q \stackrel{\text{def}}{=} 1 - p = P(X > 1)$

$$\begin{aligned}\forall n \geq 1, \quad P(X = n + 1) &= P(X > 1 \wedge X = n + 1) \\ &= P(X > 1) \cdot P(X = n + 1 | X > 1) \\ &= q \cdot P(X = n)\end{aligned}$$

Hence $P(X=2)=qp$, $P(X=3)=q^2p$,

$P(X=n)=q^{n-1}p$

Geometric

r Successes

Geometric $P(X = n) = P(\text{first success at } n\text{'th trial})$

$P(r\text{'th success at } n\text{'th trial}) = P(r - 1 \text{ successes in } n - 1 \text{ trials}) \cdot P(n\text{'th trial is success})$

$$n \geq r$$

$$b_{n-1, p}(r - 1) \quad p$$

$$= \binom{n-1}{r-1} p^{r-1} q^{n-r} p$$

$$= \binom{n-1}{r-1} p^r q^{n-r}$$

$$r = 1 \rightarrow pq^{n-1} = g_p(n)$$

Negative binomial distribution

Geometric Distribution

$$P(n) = pq^{n-1} \quad n \geq 1 \quad q = 1 - p$$

Memoryless

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

$$V(X) = \frac{q}{p^2} \quad \sigma = \frac{\sqrt{q}}{p}$$

$$P(r\text{'th success at } n\text{'th trial}) = \binom{n-1}{r-1} p^r q^{n-r}$$

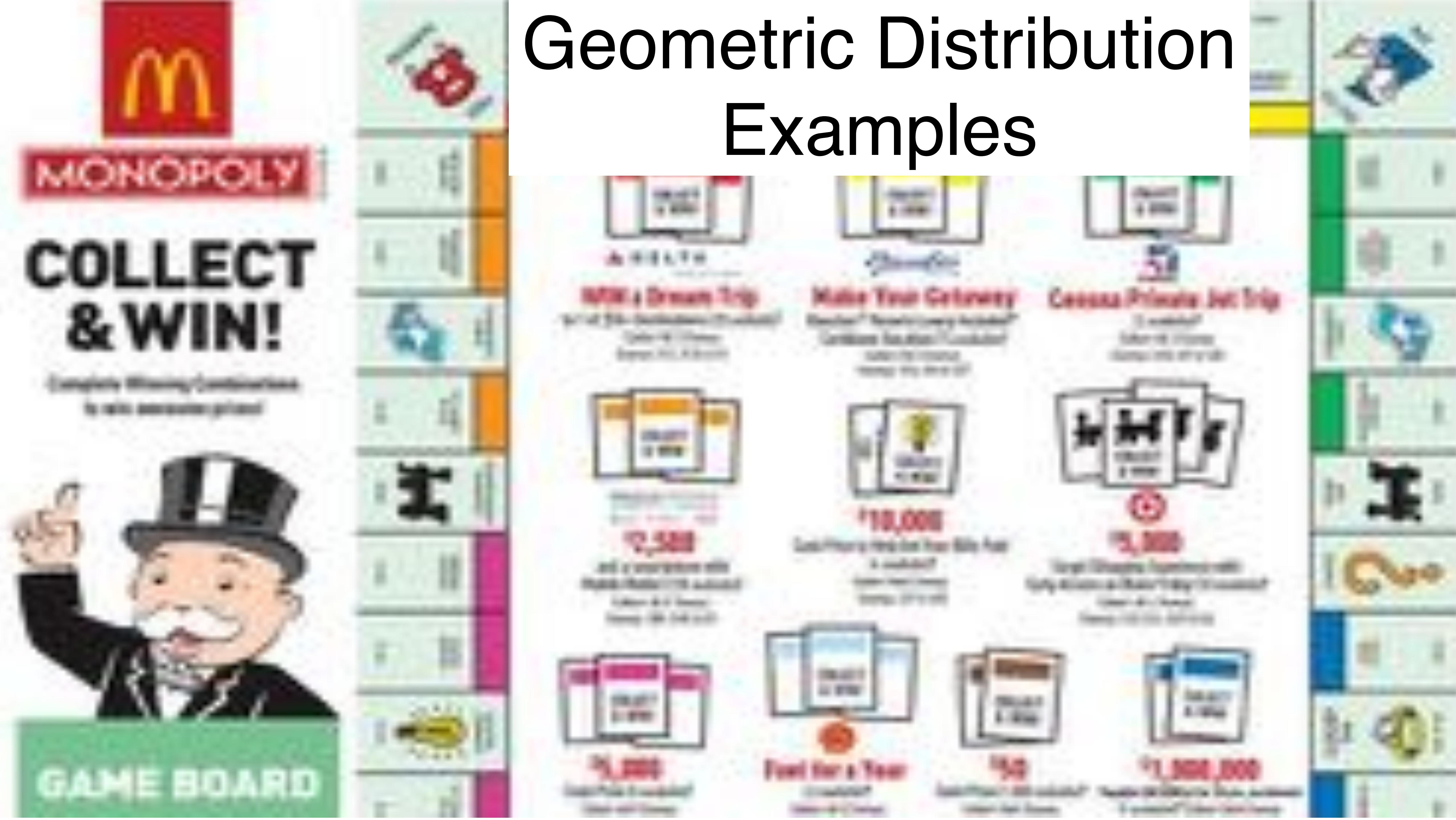
Next:



Examples



Geometric Distribution Examples



Startup Statistics

$P(\text{startup success}) = 20\%$, independent of previous attempts

Expected # startups till first success

$$X \sim G_{0.2} \quad E(X) = \frac{1}{0.2} = 5$$

Home-Grown Entrepreneur

One of first three
startups succeeds

Dad will fund up to three startups $P(\text{success})?$

$$P(X \leq 3) = F(3) = 1 - (0.8)^3 \approx 0.49$$

Cry Uncle

Even wealthier uncle funds next three startups (4,5,6)

P(success with uncle if dad's help did not suffice)?

$$\begin{aligned} P(X \in \{4, 5, 6\} | X > 3) &= P(4 | X > 3) + P(5 | X > 3) + P(6 | X > 3) \\ &= P(1) + P(2) + P(3) = P(X \leq 3) \approx 49\% \end{aligned}$$

P(success with uncle)? 1,2,3 failed but one of 4, 5, 6 succeeded

$$\begin{aligned} P(3 < X \leq 6) &= P(X > 3 \cap X \leq 6) = P(X > 3) \cdot P(x \leq 6 | x > 3) \\ &= (0.8)^3 \cdot 0.49 \approx 25\% & P(X_1, X_2, X_3 \text{ failed}) = q^3 \end{aligned}$$

$$\begin{aligned} P(3 < X \leq 6) &= F(6) - F(3) = (1 - 0.8^6) - (1 - 0.8^3) \\ &= 0.8^3 - 0.8^6 \approx 25\% \end{aligned}$$

Foreign-Born Entrepreneur

X - time to first success p=0.2

r^X - fraction of company you keep r=0.5

$$\begin{aligned} E(r^X) &= \sum_{k=1}^{\infty} r^k P(X = k) = \sum_{k=1}^{\infty} pq^{k-1} r^k = pr \sum_{i=0}^{\infty} (qr)^i \\ &= \frac{pr}{1-qr} = \frac{0.2 \cdot 0.5}{1-0.8 \cdot 0.5} = \frac{0.1}{1-0.4} = \frac{0.1}{0.6} \approx 16.67\% \end{aligned}$$



MONOPOLY

**COLLECT
& WIN!**

Complete Monopoly collection
to win amazing prizes



GAME BOARD

Coupon Collector Problem

The image shows a standard Monopoly board with several coupon cards placed on its spaces. The cards represent different prizes and requirements for collecting them. The board includes the following details:

- Properties:** Boardwalk, Water Works, Park Place, Mayfield, Baltic Avenue, Atlantic Avenue, Marvin Gardens, New York Avenue, St. Charles Place, St. Charles Avenue, Illinois Avenue, Pennsylvania Avenue, Park Place, and Boardwalk.
- Coupons:** There are four sets of three coupons each, representing different categories:
 - McDonald's:** Includes "Free Large Soft Drink" and "Free Large Fries".
 - Starbucks:** Includes "Free Medium Latte" and "Free Medium Cappuccino".
 - Cinemark:** Includes "Free Large Popcorn" and "Free Large Soda".
 - Gasoline:** Includes "Free Gasoline" and "Free Gasoline".
- Prizes:** There are three large boxes representing major prizes:
 - McDonald's Grand Prize:** "Get a Dream Trip" (with details about a trip to Disney World).
 - Starbucks Grand Prize:** "Get a Dream Trip" (with details about a trip to Disney World).
 - Cinemark Grand Prize:** "Get a Dream Trip" (with details about a trip to Disney World).
- Other Labels:** "100,000" is printed near the Cinemark space, and "1,000,000" is printed near the Gasoline space.

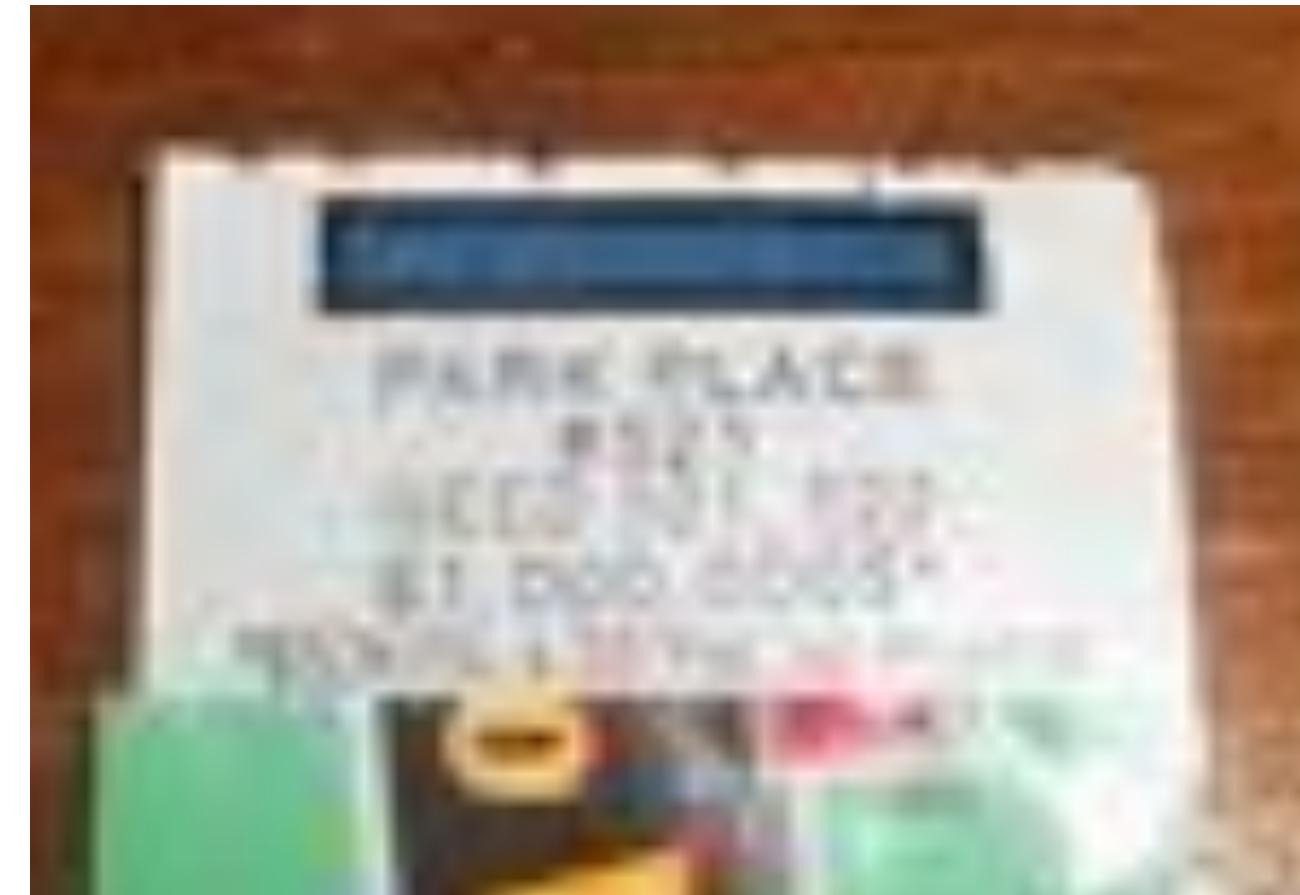
Pre



n coupons

Each item contains one coupon selected uniformly

Collect all coupons, get a prize

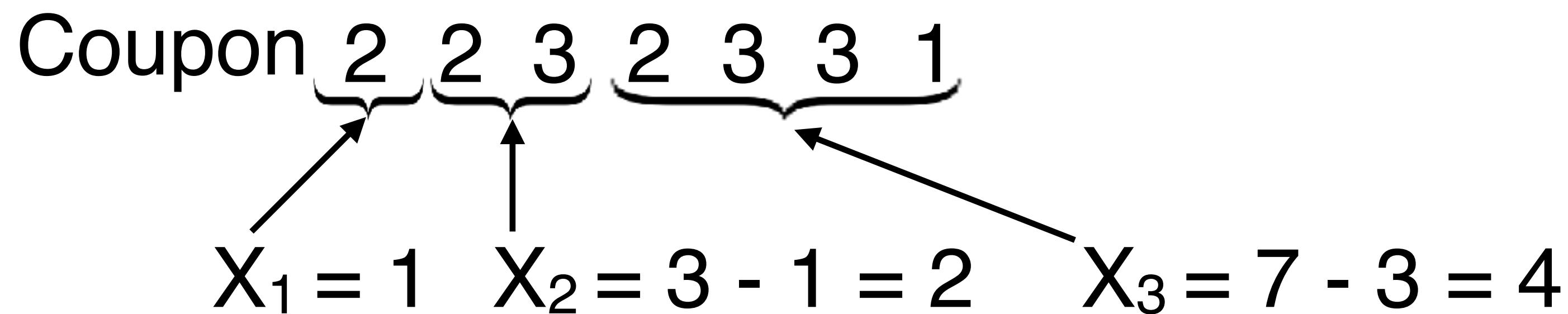


How many items need to buy to collect all?

Expectation

X - # items to collect all coupons

$n = 3$ Items 1 2 3 4 5 6 7 $X = 7$ EX?



X_i - # items to get i^{th} coupon after getting $i - 1$ coupons

$$X = X_1 + X_2 + X_3$$

$$7 = 1 + 2 + 4$$

$$\left. \begin{array}{l} X_1 = 1 \\ X_2 \sim G_{2/3} \\ X_3 \sim G_{1/3} \end{array} \right\} \perp\!\!\!\perp$$

General n

$$X_i \sim G\left(\frac{n-(i-1)}{n}\right) = G\left(\frac{n-i+1}{n}\right)$$

$$EX_i = \frac{n}{n-i+1}$$

$$X = \sum_{i=1}^n X_i$$

$$EX = \sum_{i=1}^n EX_i = \sum_{i=1}^n \frac{n}{n-i+1} = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \dots + \frac{n}{1}$$

$$= n\left(\frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{n}\right) = nH_n \approx n \ln n + 0.577n$$

$$\text{Harmonic Sum } H_n = \frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{n} \rightarrow \ln n + 0.577\dots$$

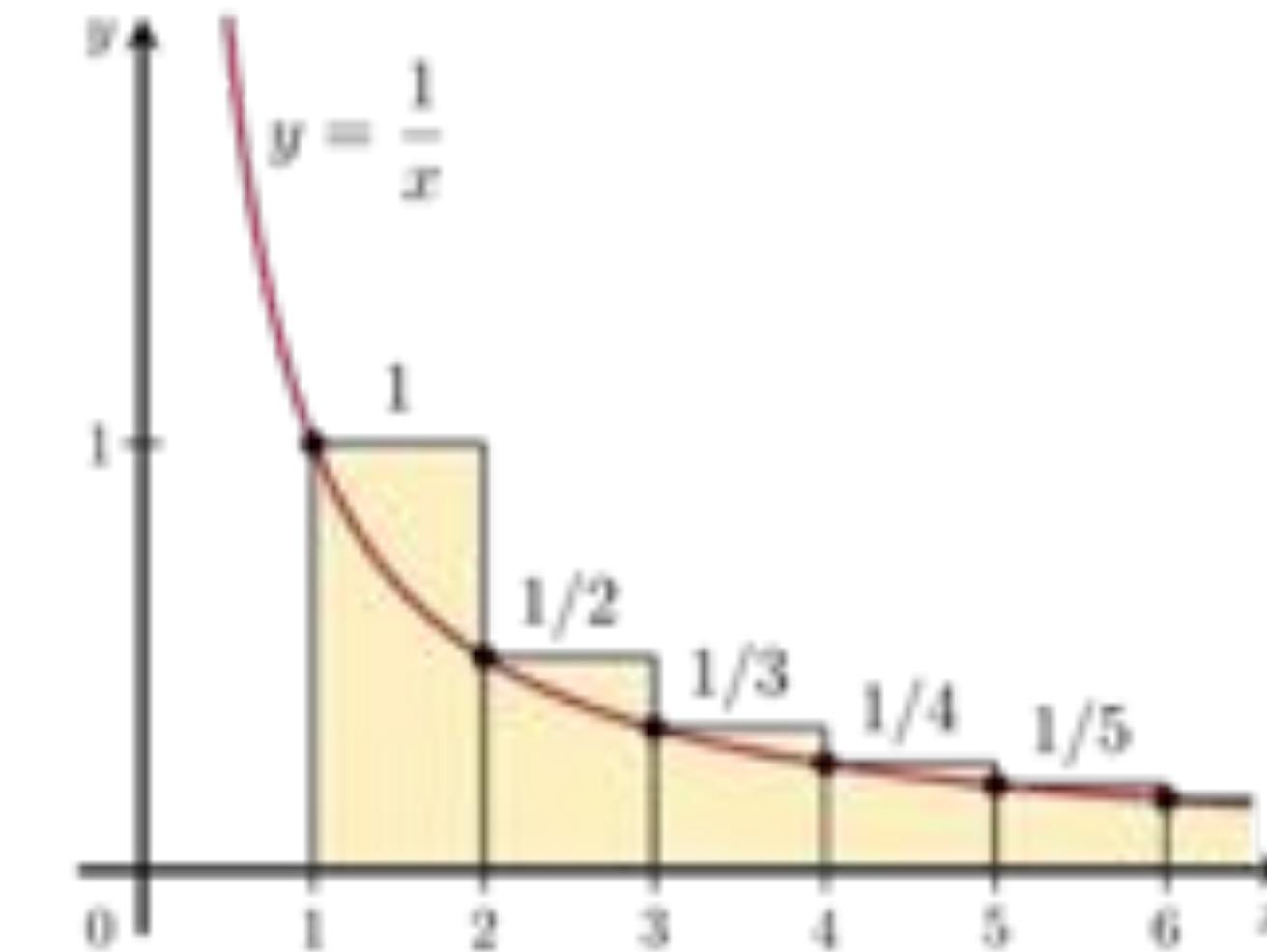
Harmonic Sum

$$H_n > \int_1^{n+1} \frac{1}{x} dx$$

$$= \ln x \Big|_1^{n+1} = \ln(n+1)$$

$$H_n \leq 1 + \int_1^n \frac{1}{x} dx = 1 + \ln x \Big|_1^n = 1 + \ln n$$

$$H_n \rightarrow \ln n + 0.577\dots$$



Variance

$$X \sim G(P)$$

$$V(X) = \frac{1-p}{p^2} \leq \frac{1}{p^2}$$

$$V(X) = V\left(\sum_{i=1}^n X_i\right)$$

$$\stackrel{\textcircled{\text{II}}}{=} \sum_{i=1}^n V(X_i)$$

$$\leq \sum_{i=1}^n \frac{1}{(\frac{n-i+1}{n})^2}$$

$$= n^2 \left(\frac{1}{n^2} + \frac{1}{(n-1)^2} + \dots + \frac{1}{1^2} \right)$$

$$\leq \frac{\pi^2}{6} n^2$$

$$\sigma \leq \frac{\pi}{\sqrt{6}} n$$

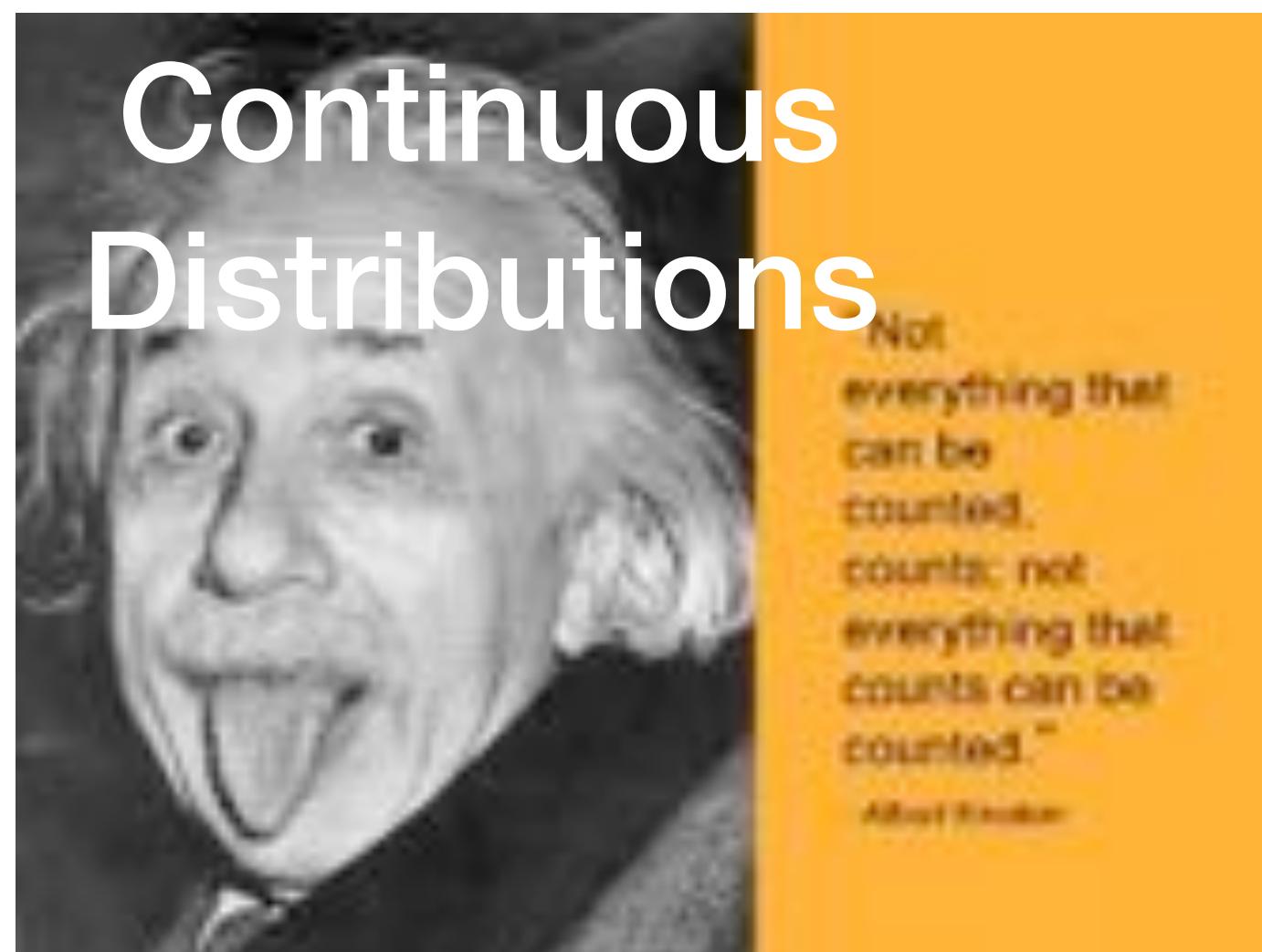
Summary

Geometric-distribution examples

Coupon collector problem

Discrete distribution families

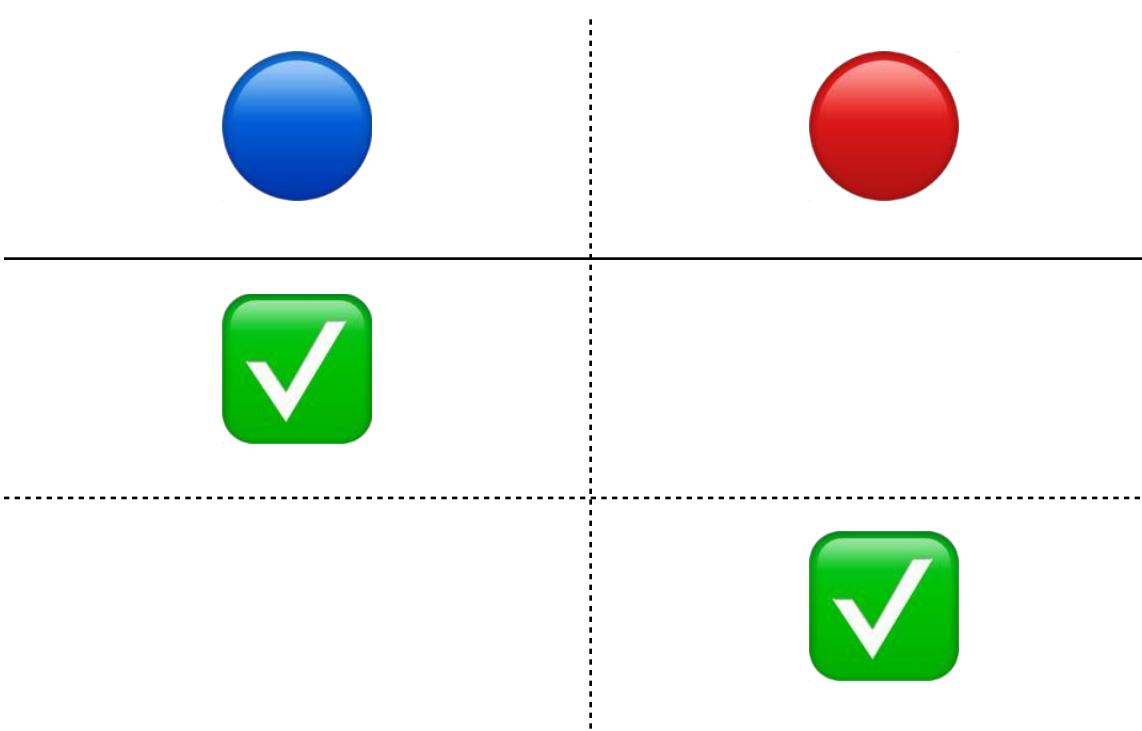
Bernoulli, Binomial, Poisson, Geometric



Hypergeometric Distribution

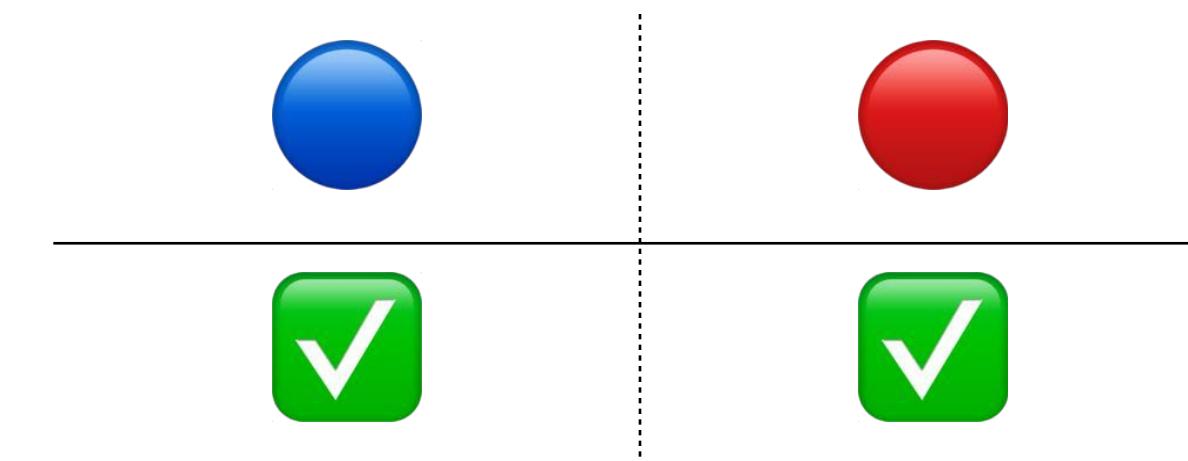
$$b = r = 1 \quad n = 1$$

k	$P(k)$
0	$\frac{1}{2}$
1	$\frac{1}{2}$



$$b = r = 1 \quad n = 2$$

k	$P(k)$
0	0
1	1
2	0



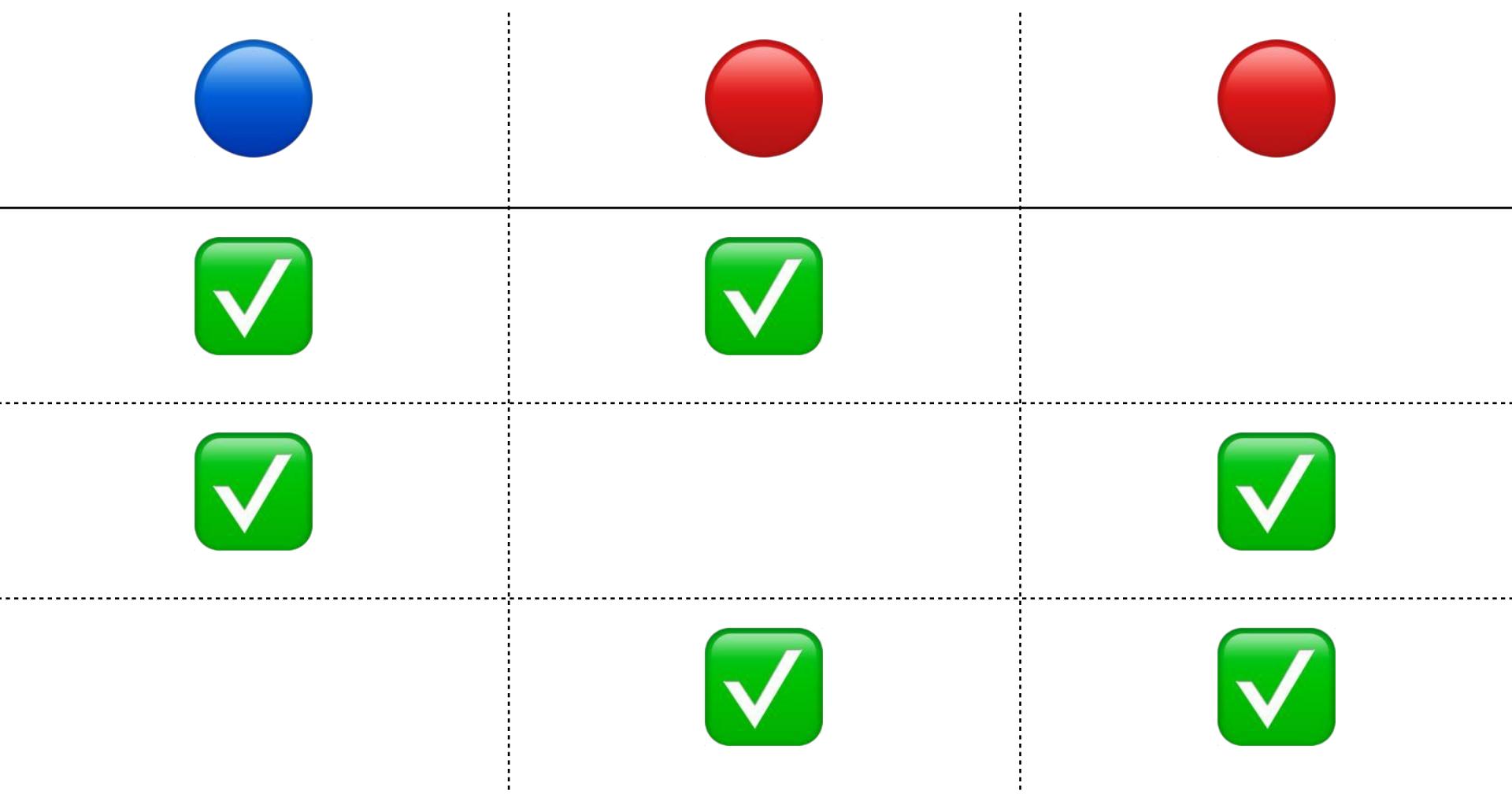
$$P(0) = \frac{\binom{1}{0} \cdot \binom{1}{2}}{\binom{2}{2}} = 0$$

$$P(1) = \frac{\binom{1}{1} \cdot \binom{1}{1}}{\binom{2}{2}} = \frac{1 \cdot 1}{1} = 1$$

$$P(2) = 0$$

$$b = 1 \quad r = 2 \quad n = 2$$

k	$P(k)$
0	$\frac{1}{3}$
1	$\frac{2}{3}$
2	0



$$P(k \text{ blue balls}) = \frac{\binom{b}{k} \cdot \binom{r}{n-k}}{\binom{b+r}{n}}$$

Hypergeometric Distribution

b blue balls r red balls

Pick $n \geq 0$ balls

$$|\Omega| = \binom{b+r}{n}$$

If k blue balls, then $n-k$ red balls

$$P(k \text{ blue balls}) = \frac{\binom{b}{k} \cdot \binom{r}{n-k}}{\binom{b+r}{n}}$$

Recall

$$\sum_{k=0}^n \binom{b}{k} \binom{r}{n-k} = \binom{b+r}{n}$$

$$\sum_{k=0}^n P(k) = \frac{\sum_{k=0}^n \binom{b}{k} \binom{r}{n-k}}{\binom{b+r}{n}} = \frac{\binom{b+r}{n}}{\binom{b+r}{b}} = 1$$

Socks

Box contains 4 white, 3 black socks Choose 5

Find probability of 0,1,2,3,4,5 white

Total # choices $\binom{7}{5} = \binom{7}{2} = \frac{7 \cdot 6}{2} = 21$

0 white \rightarrow 5 black \rightarrow impossible

1 white \rightarrow 4 black \rightarrow impossible

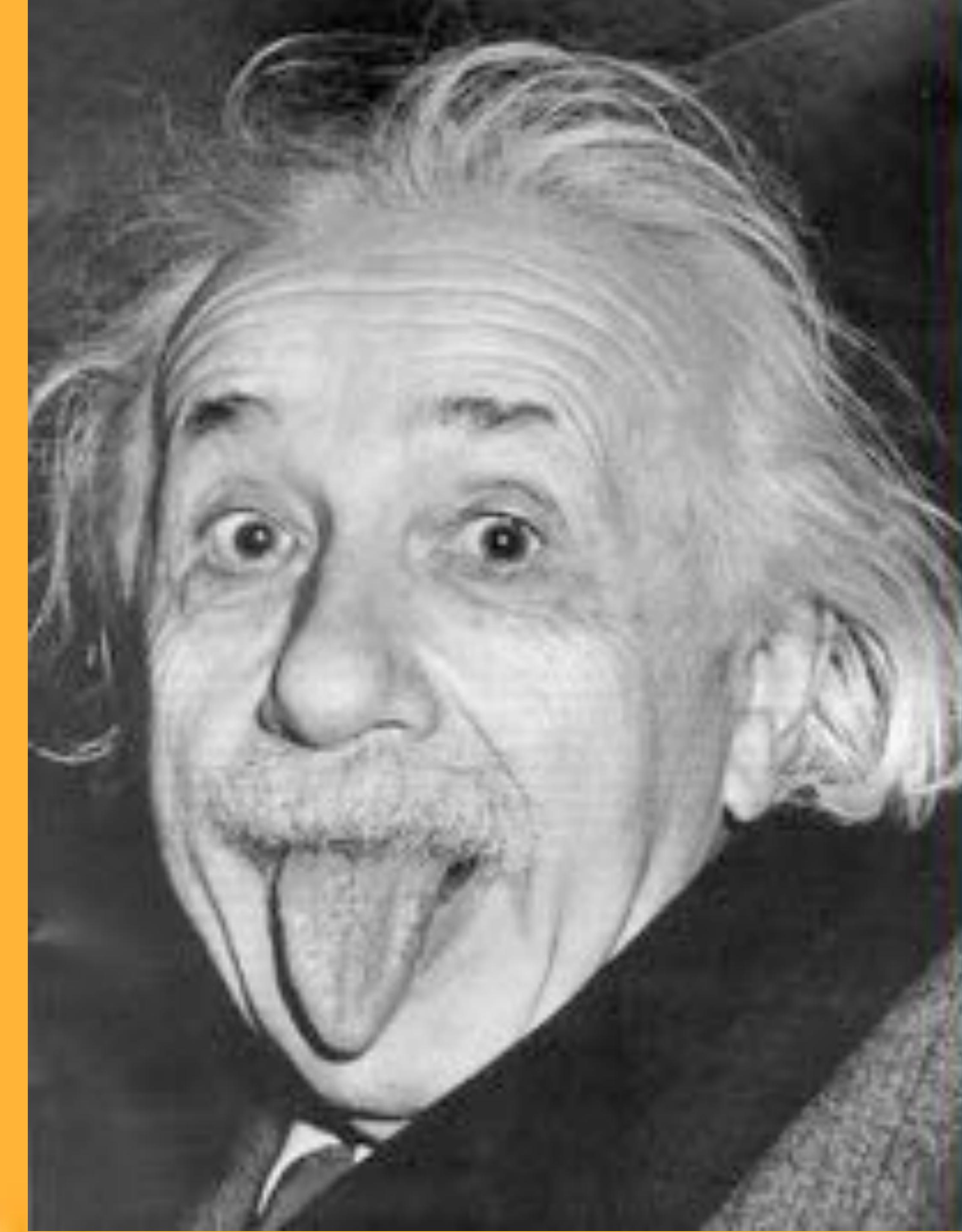
2 white \rightarrow 3 black $\rightarrow \frac{\binom{3}{3} \binom{4}{2}}{21} = \frac{6}{21}$

3 white \rightarrow 2 black $\rightarrow \frac{\binom{3}{2} \binom{4}{3}}{21} = \frac{12}{21}$

4 white \rightarrow 1 black $\rightarrow \frac{\binom{3}{1} \binom{4}{4}}{21} = \frac{3}{21}$

5 white \rightarrow impossible

Continuous Distributions



"Not everything that can be counted, counts; not everything that counts can be counted."

- Albert Einstein

Discrete to Continuous

Discrete distributions: Countable # values (finite or countably-infinite)

Continuous distributions: Uncountable # values, intervals

Why Continuous

Anything physics

Time

flight

delivery

disease

life

Space

height

storm area

Mass

pet

cookie

Temperature

air

body

Nearly continuous variables

Cost

stock

house

pork bellies

Rates

interest

exchange

unemployment

Probability Density Function

Replaces the discrete pmf

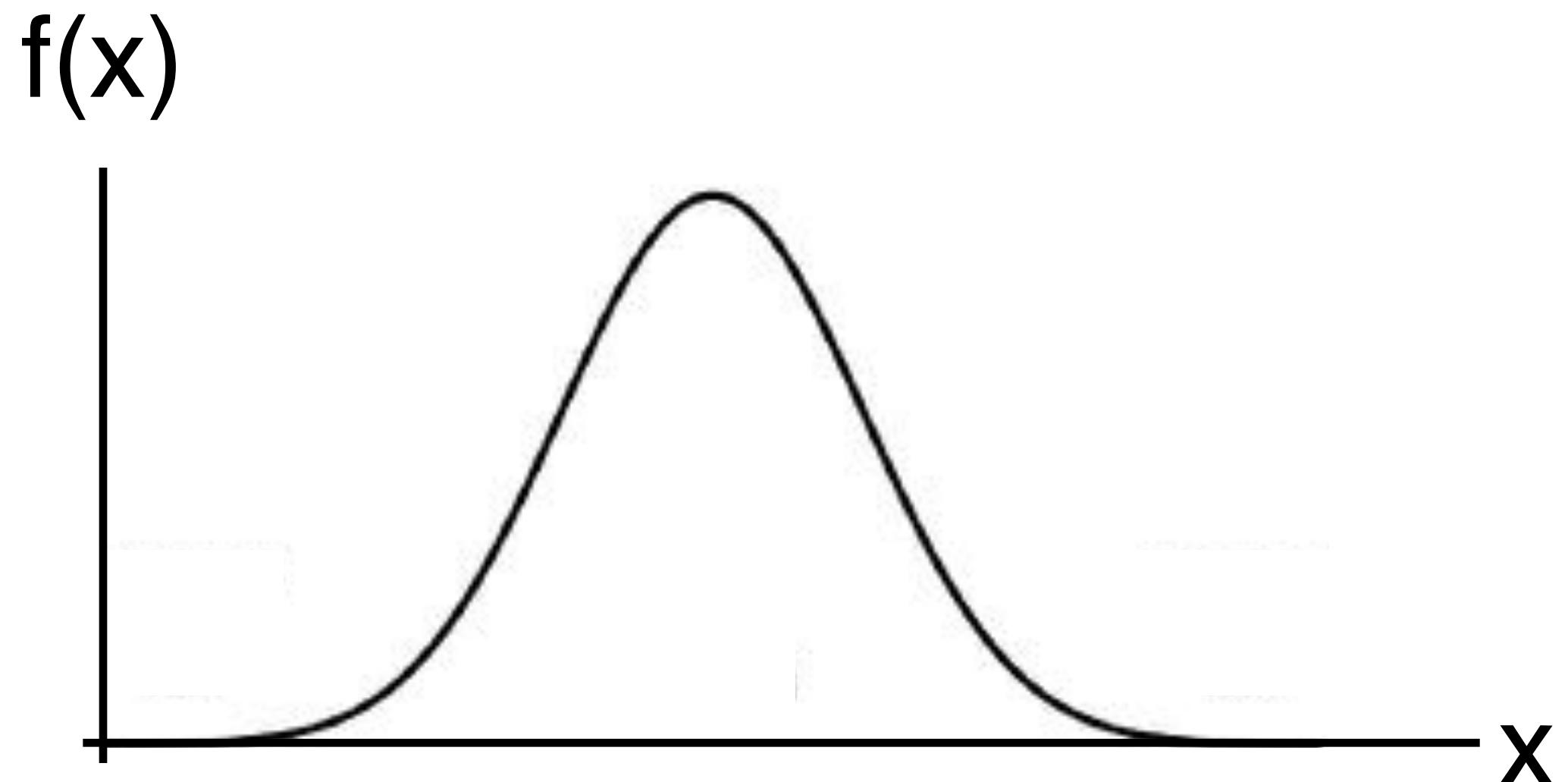
$$f(x) \geq 0$$

relative likelihood of x

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

area under curve

(area)

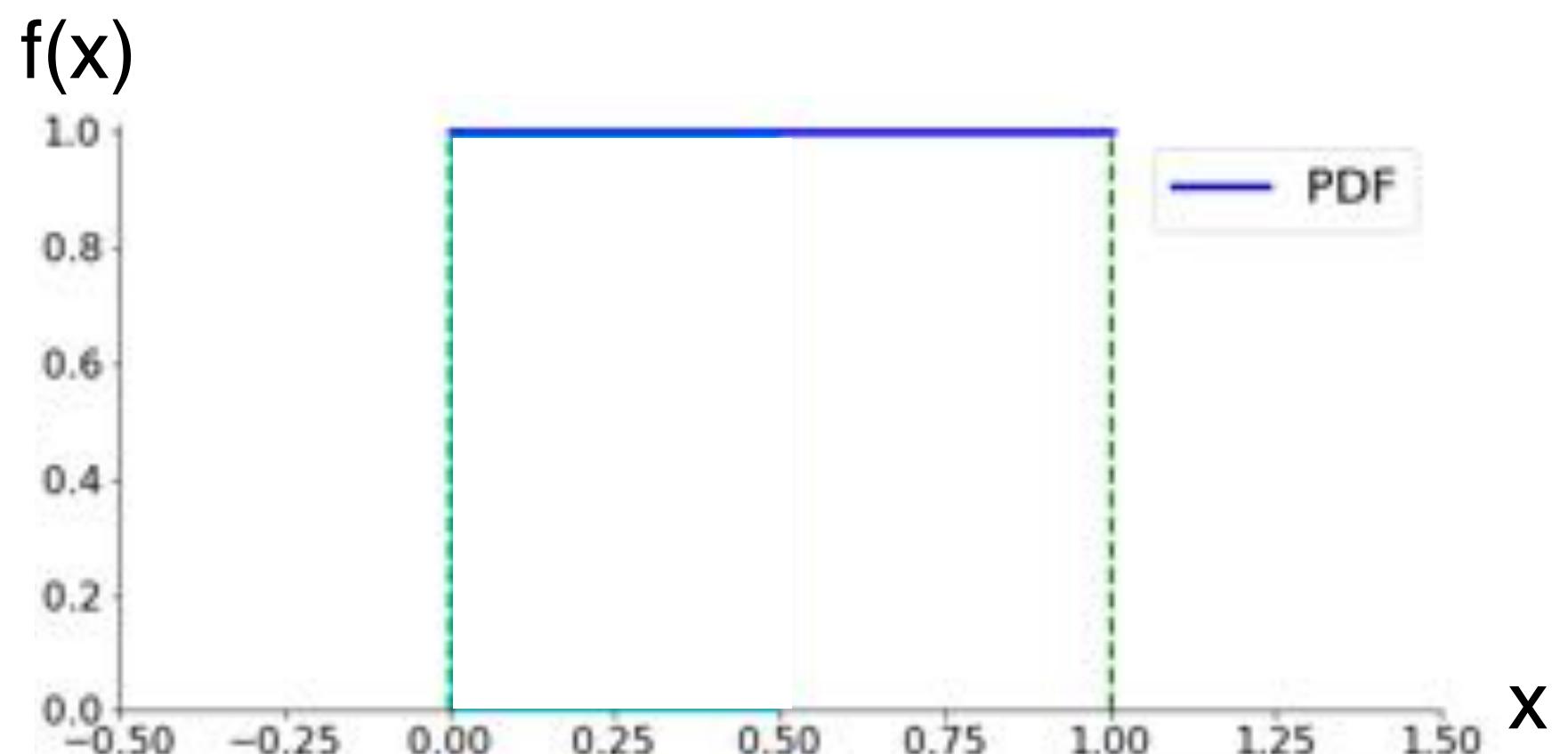


Comparison to Discrete

	Discrete	Continuous
Probability function	mass (pmf)	density (pdf)
≥ 0	$p(x) \geq 0$	$f(x) \geq 0$
$\sum = 1$	$\sum_x p(x) = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$

Uniform

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$



Will it \int ? Area = $1 \cdot 1 = 1$

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^1 1dx = x \Big|_0^1 = 1$$

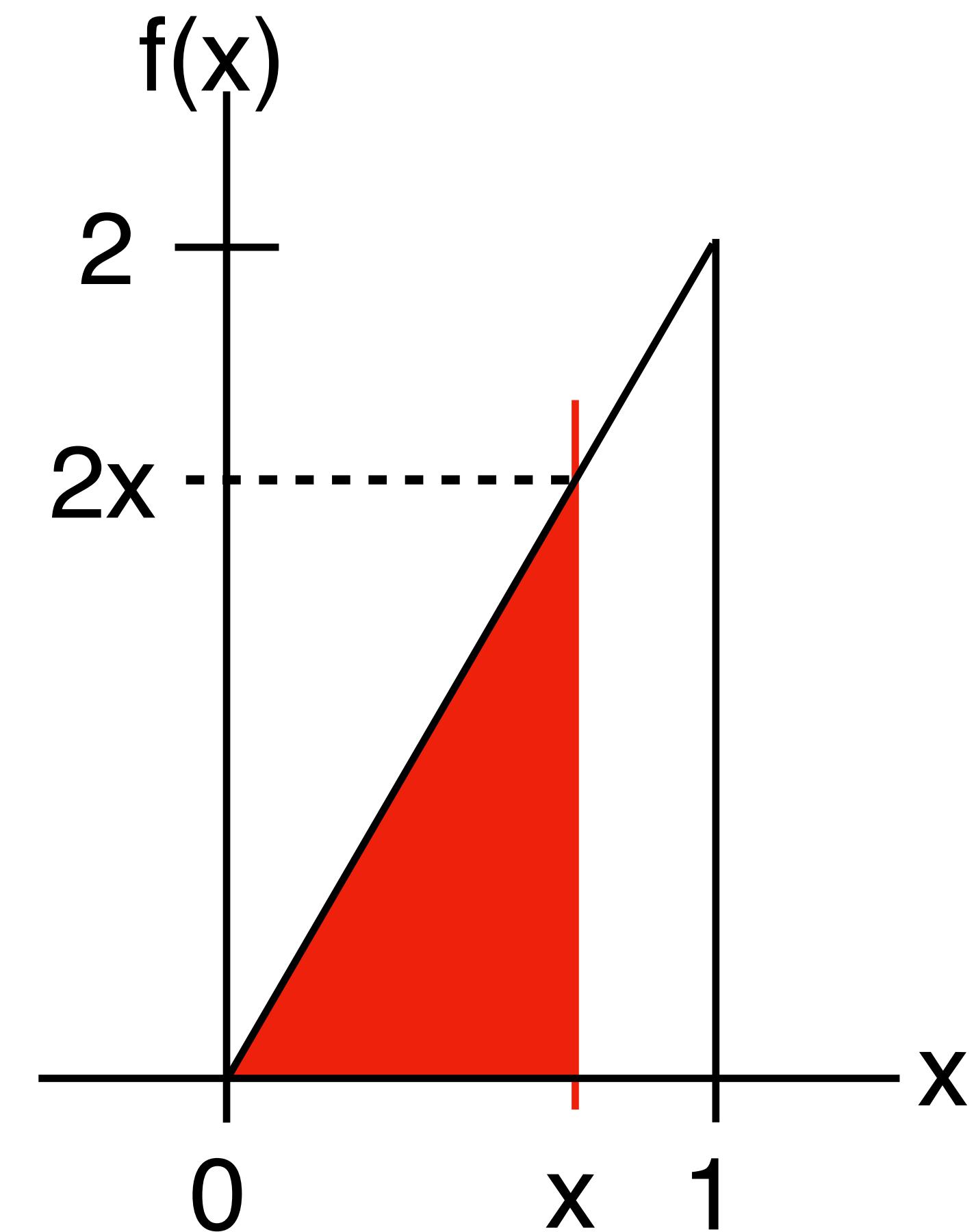
Triangle

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Will it \int ?

$$\text{Area} = 2 \cdot 1 \cdot \frac{1}{2} = 1$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^1 2xdx = x^2 \Big|_0^1 = 1 - 0 = 1$$



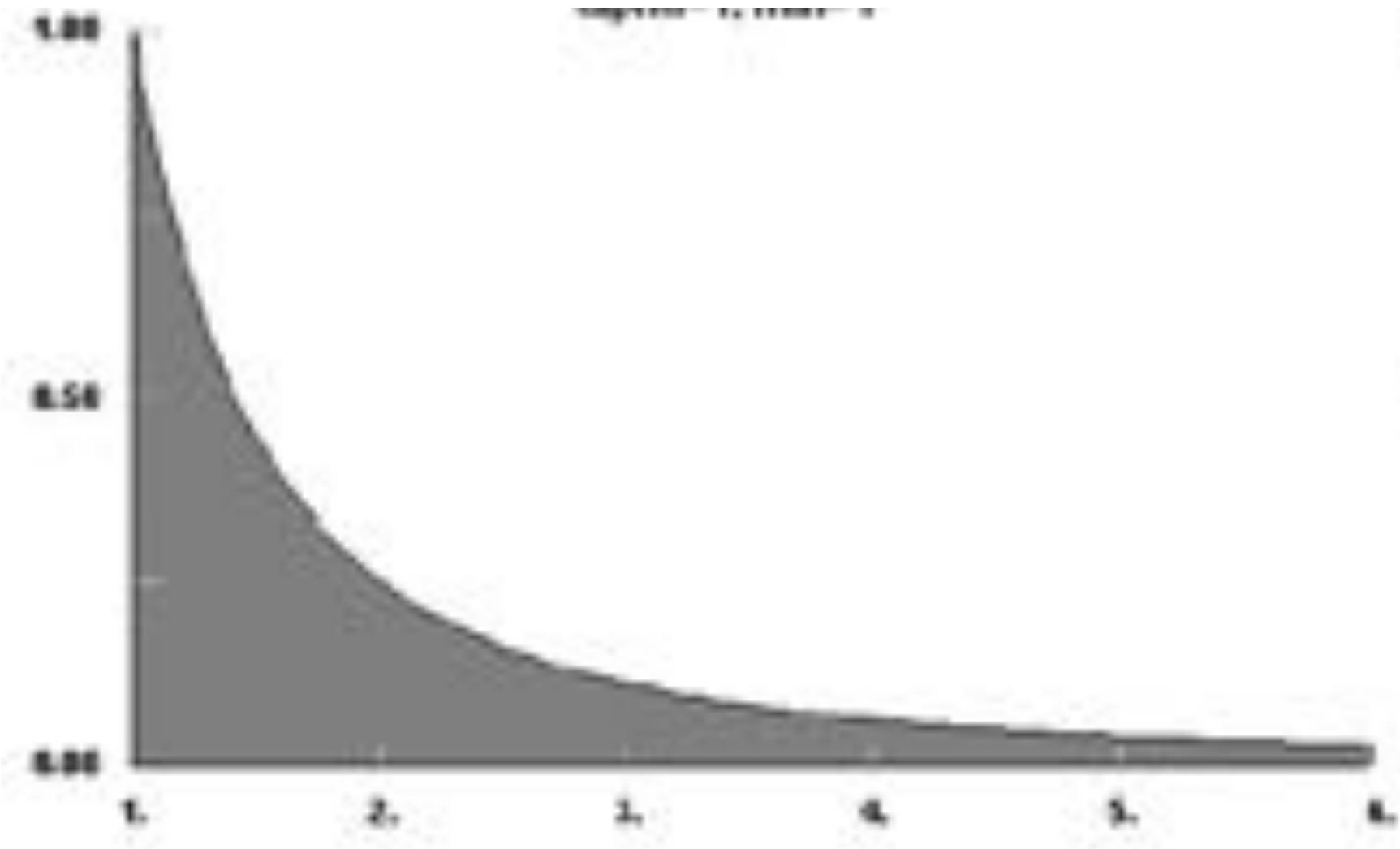
Infinite Support

Power law

$$f(x) = \begin{cases} \frac{1}{x^2} & x \geq 1 \\ 0 & x < 1 \end{cases}$$

Will it Σ ?

$$\int_{-\infty}^{\infty} f(x)dx = \int_1^{\infty} \frac{1}{u^2} du = \left. -\frac{1}{u} \right|_1^{\infty} = 1$$



Event Probability

	Discrete	Continuous
P(A)	$\sum_{x \in A} p(x)$	$\int_{x \in A} f(x) dx$

Typically interested in interval probability $P(a \leq X \leq b)$

Area between a and b

$P(X \leq b) - P(X \leq a)$

Cumulative distribution function

Cumulative Distribution Function (CDF)

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

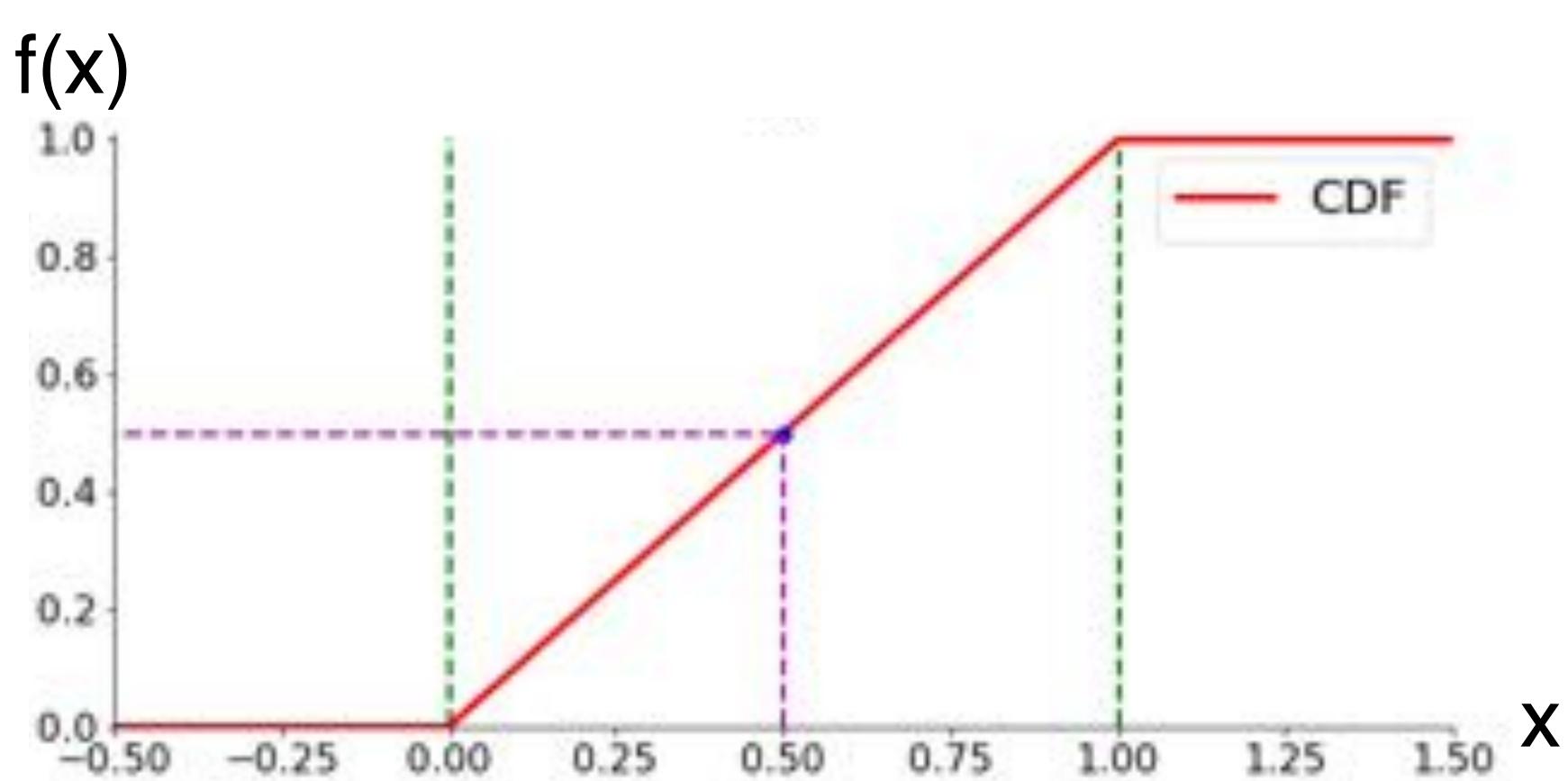
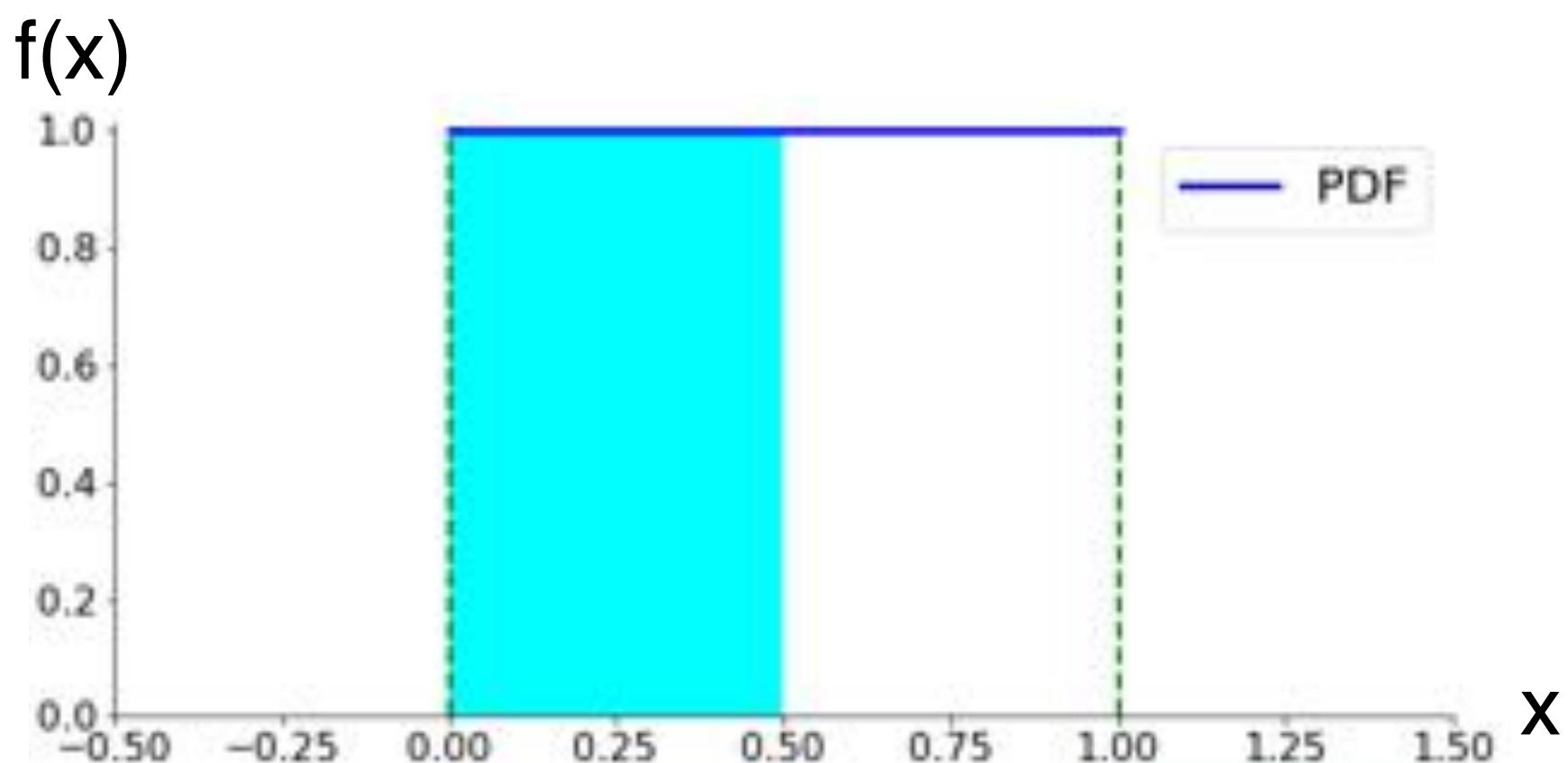
	Discrete	Continuous
PF → CDF	$\sum_{u \leq x} p(u)$	$\int_{-\infty}^x f(u) du$
CDF → PF	$p(x) = F(x) - F(x^*)$	$f(x) = F'(x)$

x^* - element preceding x

Uniform

$$F(x) = \int_{-\infty}^x f(u)du = \begin{cases} 0 & x \leq 0 \\ \int_0^x 1 du = u \Big|_0^x = x & 0 \leq x \leq 1 \\ 1 & 1 \leq x \end{cases}$$

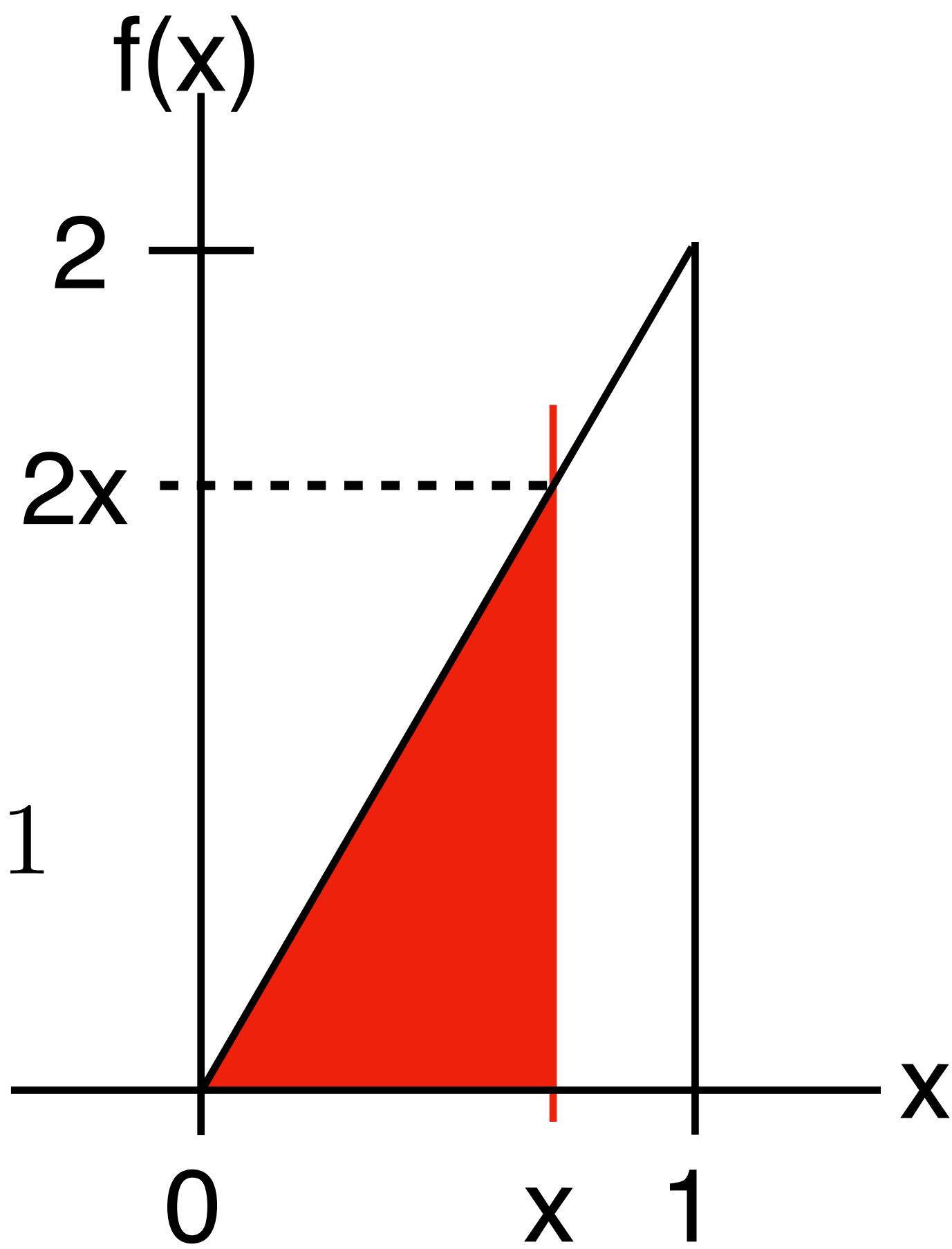
$$F'(x) = \begin{cases} (0)' = 0 & x \leq 0 \\ (x)' = 1 & 0 < x < 1 \\ (1)' = 0 & 1 < x \end{cases}$$



Triangle

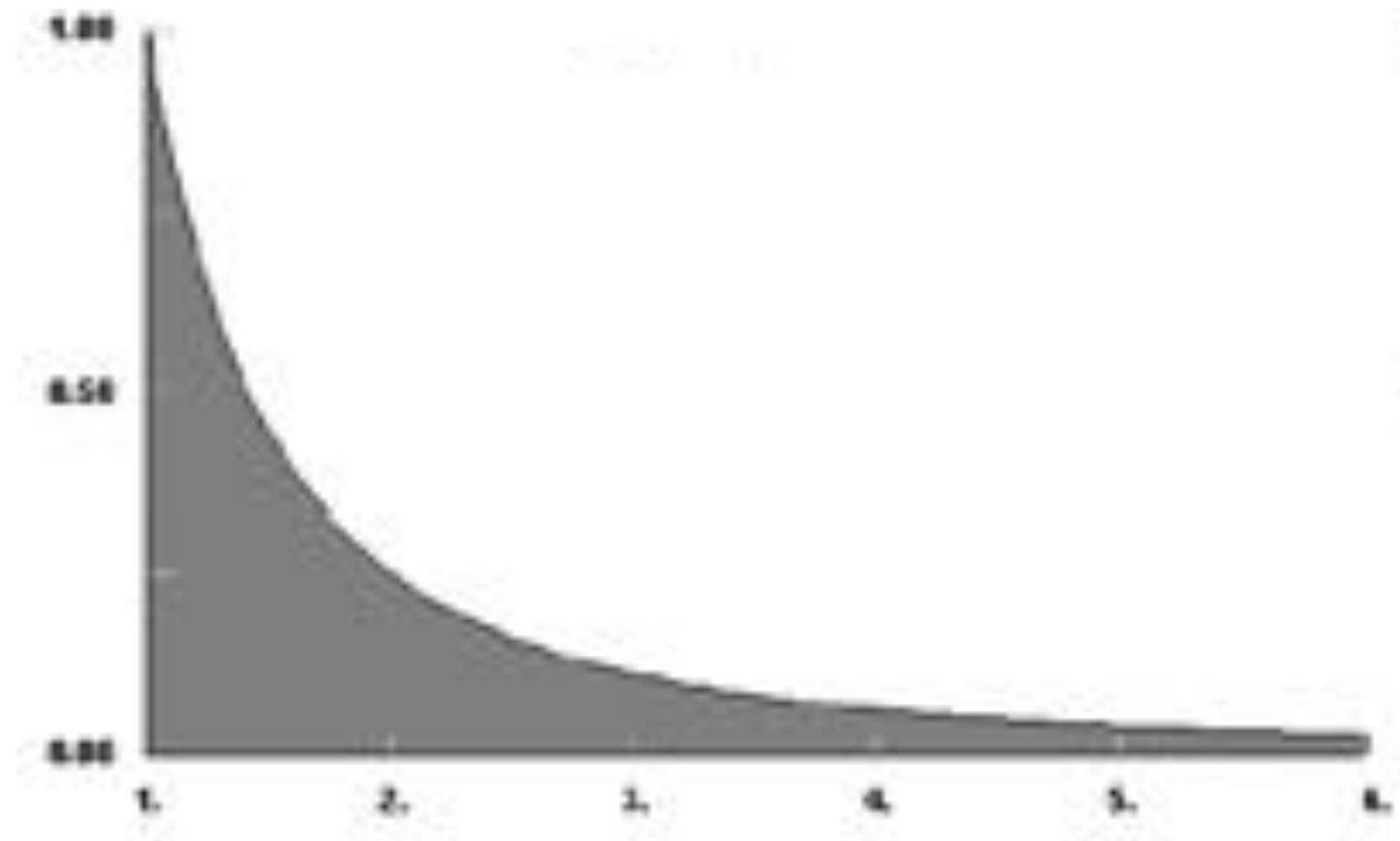
$$F(x) = \int_{-\infty}^x f(u)du = \begin{cases} 0 & x \leq 0 \\ \int_0^x 2udu = u^2 \Big|_0^x = x^2 & 0 \leq x \leq 1 \\ 1 & 1 \leq x \end{cases}$$

$$F'(x) = \begin{cases} (0)' = 0 & x < 0 \\ (x^2)' = 2x & 0 < x < 1 \\ (1)' = 0 & 1 < x \end{cases}$$



Infinite Support

$$F(x) = \begin{cases} 0 & x \leq 1 \\ \int_1^x \frac{1}{u^2} du = \frac{-1}{u} \Big|_1^x = 1 - \frac{1}{x} & x \geq 1 \end{cases}$$



$$F'(x) = \begin{cases} (0)' = 0 & x < 1 \\ \left(1 - \frac{1}{x}\right)' = \frac{1}{x^2} & x > 1 \end{cases}$$

Properties of the CDF

$F(x) = \text{integral}$

Nondecreasing

$F(-\infty)=0$

$F(\infty)=1$

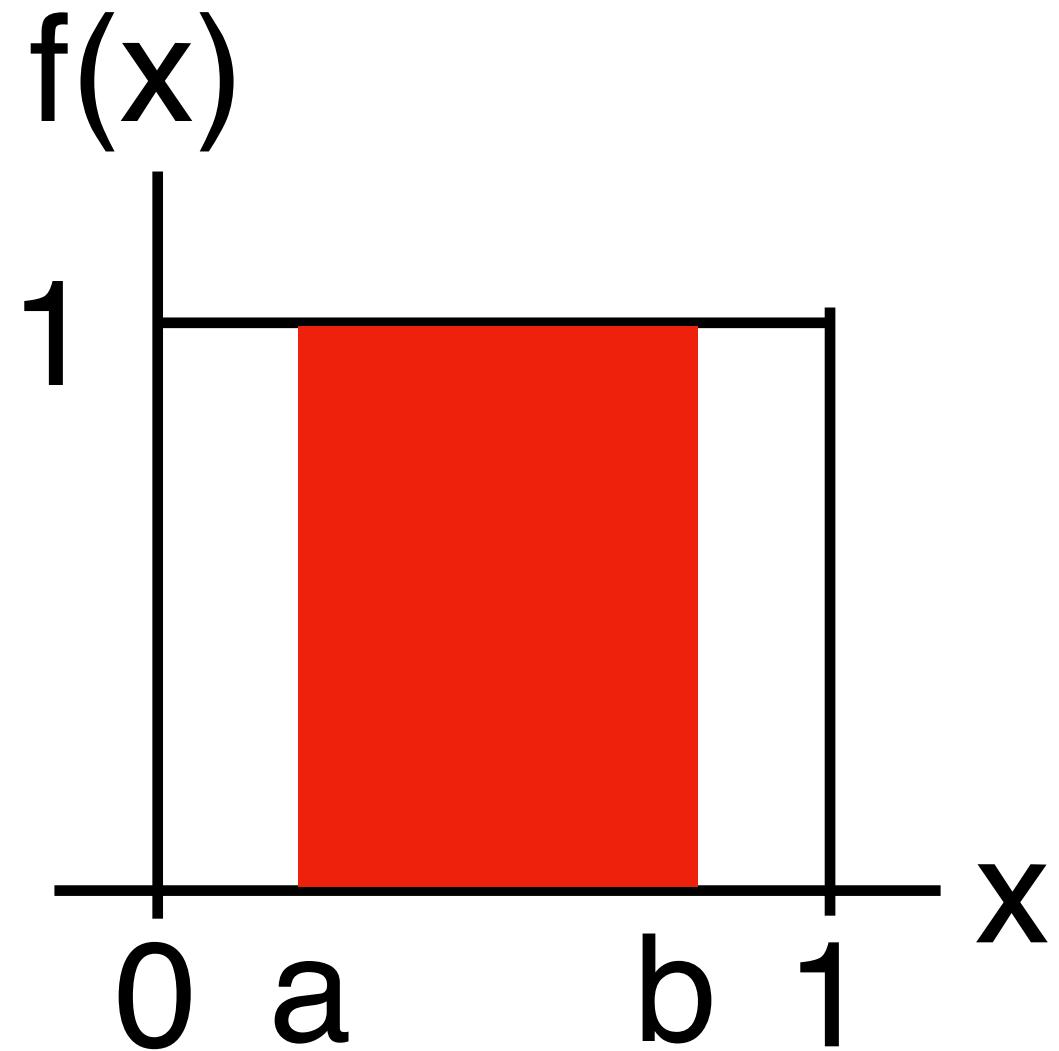
Continuous

Interval Probabilities

Examples

Uniform

$$0 \leq a \leq b \leq 1$$
$$P(a \leq X \leq b) = \begin{cases} \text{Area} = (b - a) \cdot 1 = b - a \\ \int_a^b f(x)dx = \int_a^b 1 dx = x \Big|_a^b = b - a \\ F(b) - F(a) = b - a \end{cases}$$



$$\begin{aligned} P(0.6 \leq X \leq 1.3) &= P(0.6 \leq X \leq 1) = 0.4 \\ &= F(1.3) - F(0.6) = 1 - 0.6 = 0.4 \end{aligned}$$

Power law

$$1 \leq a \leq b \quad P(a \leq X \leq b) = F(b) - F(a) = \left(1 - \frac{1}{b}\right) - \left(1 - \frac{1}{a}\right) = \frac{1}{a} - \frac{1}{b}$$

Differences

Discrete	Continuous
$p(x) \leq 1$	$f(x)$ can be > 1
Generally $p(x) \neq 0$	$p(x) = 0$
Generally $P(X \leq a) \neq P(X < a)$	$P(X \leq a) = P(X < a) = F(a)$
	$P(X \geq a) = P(X > a) = 1 - F(a)$
	$P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a)$

Expectation

	Discrete	Continuous
EX	$\sum x \cdot p(x)$	$\int_{-\infty}^{\infty} xf(x)dx$

As discrete:

Average of many samples

Properties

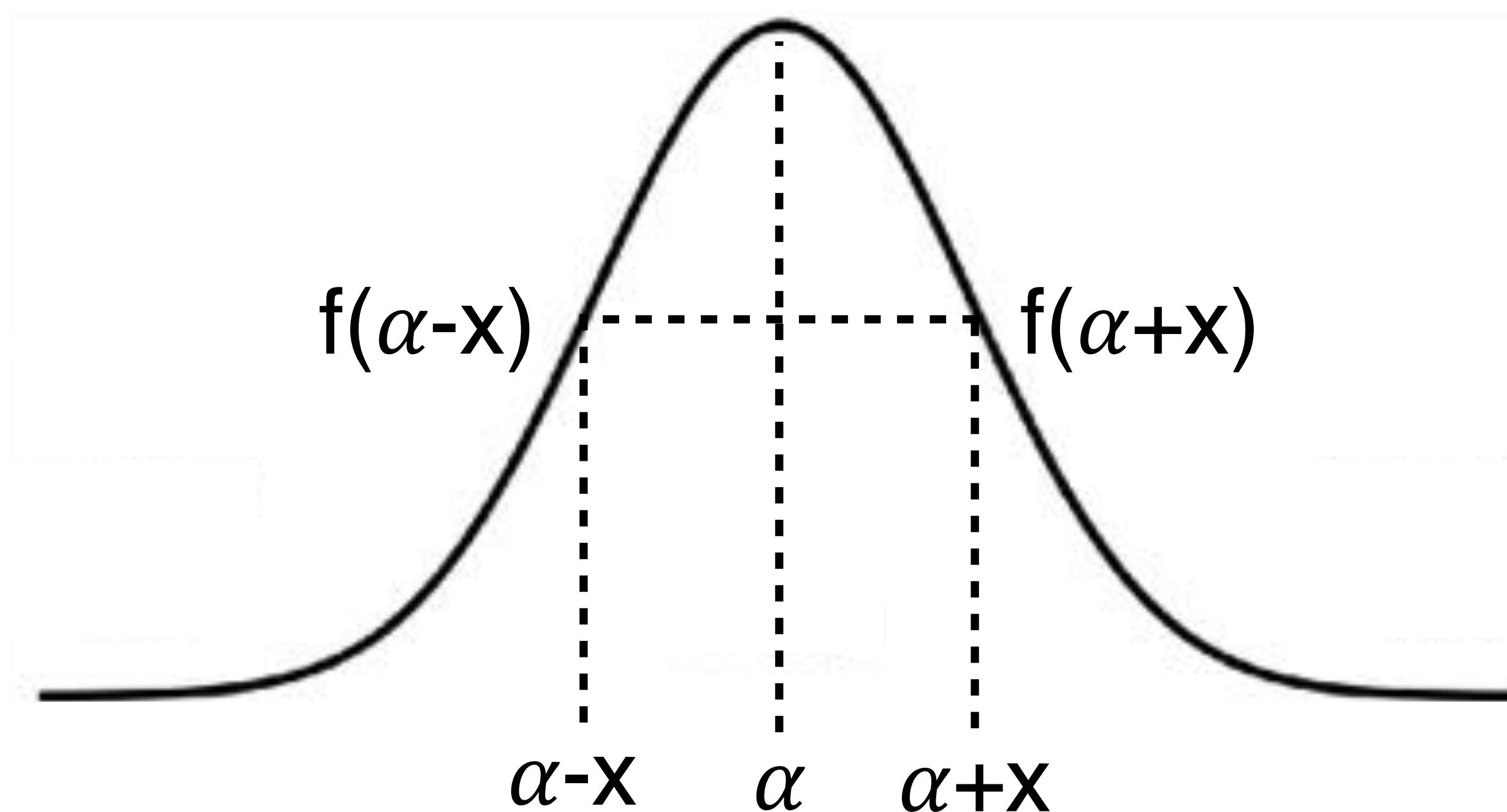
Support set = $[a, b]$

$$a \leq EX \leq b$$

Symmetry

If for some α , $f(\alpha+x) = f(\alpha-x)$ for all x

then $EX = \alpha$



Examples

Uniform $EX = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 x 1 dx = \frac{x^2}{2} \Big|_0^1 = \frac{1}{2}$

Triangle $EX = \int_0^1 x 2x dx = \frac{2x^3}{3} \Big|_0^1 = \frac{2}{3}$

Power law $EX = \int_1^{\infty} x \frac{1}{x^2} dx = \int_1^{\infty} \frac{1}{x} dx = \ln x \Big|_1^{\infty} = \infty$

Later: power laws with finite expectation

Variance

	Discrete	Continuous
$V(X) \triangleq E(X - \mu)^2$	$\sum_x p(x)(x - \mu)^2$	$\int_{-\infty}^{\infty} f(x)(x - \mu)^2 dx$
		$E(X^2) - \mu^2$

As for discrete

$$\begin{aligned} E(X - \mu)^2 &= \int (x - \mu)^2 f(x) dx \\ &= \int (x^2 - 2x\mu + \mu^2) f(x) dx \\ &= \int x^2 f(x) dx - 2\mu \int x f(x) dx + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

$$\sigma = \sqrt{V(X)}$$

Examples

Uniform

$$EX = \frac{1}{2}$$

$$E(X^2) = \int_0^1 x^2 \cdot 1 \, dx = \frac{x^3}{3} \Big|_0^1 = \frac{1}{3}$$

$$V(X) = E(X^2) - (EX)^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

$$\sigma = \frac{1}{\sqrt{12}} = \frac{1}{2\sqrt{3}}$$

Triangle

$$EX = \frac{2}{3}$$

$$EX^2 = \int_0^1 x^2 \cdot 2x \, dx = \frac{2}{4}x^4 \Big|_0^1 = \frac{1}{2}$$

$$V(X) = E(X^2) - (EX)^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{9-8}{18} = \frac{1}{18}$$

$$\sigma = \frac{1}{\sqrt{18}} = \frac{1}{3\sqrt{2}}$$

Discrete vs. Continuous

	Discrete	Continuous
Prob. Fun.	pmf - p	pdf - f
≥ 0	$p(x) \geq 0$	$f(x) \geq 0$
$\sum = 1$	$\sum p(x) = 1$	$\int f(x)dx = 1$
P(A)	$\sum_{x \in A} p(x)$	$\int_{x \in A} f(x)dx$
F(X)	$\sum_{u \leq x} p(u)$	$\int_{-\infty}^x f(u)du$
$\mu = E(X)$	$\sum xp(x)$	$\int xf(x)dx$
V(X)	$\sum (x - \mu)^2 p(x)$	$\int (x - \mu)^2 f(x)dx$

$p(x) \leq 1$ f can be larger

$$f(x) = F'(x)$$

$$P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a)$$

A cartoon illustration of a man with a mustache and a green shirt looking through a blue telescope. He has a white and blue striped hat and is wearing an orange vest over his shirt.

Functions of Continuous Random Variables