Geometric distributions

Back to Basics

Independent B_p coin flips

$p(1) = p$

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

Two Derived Distributions

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

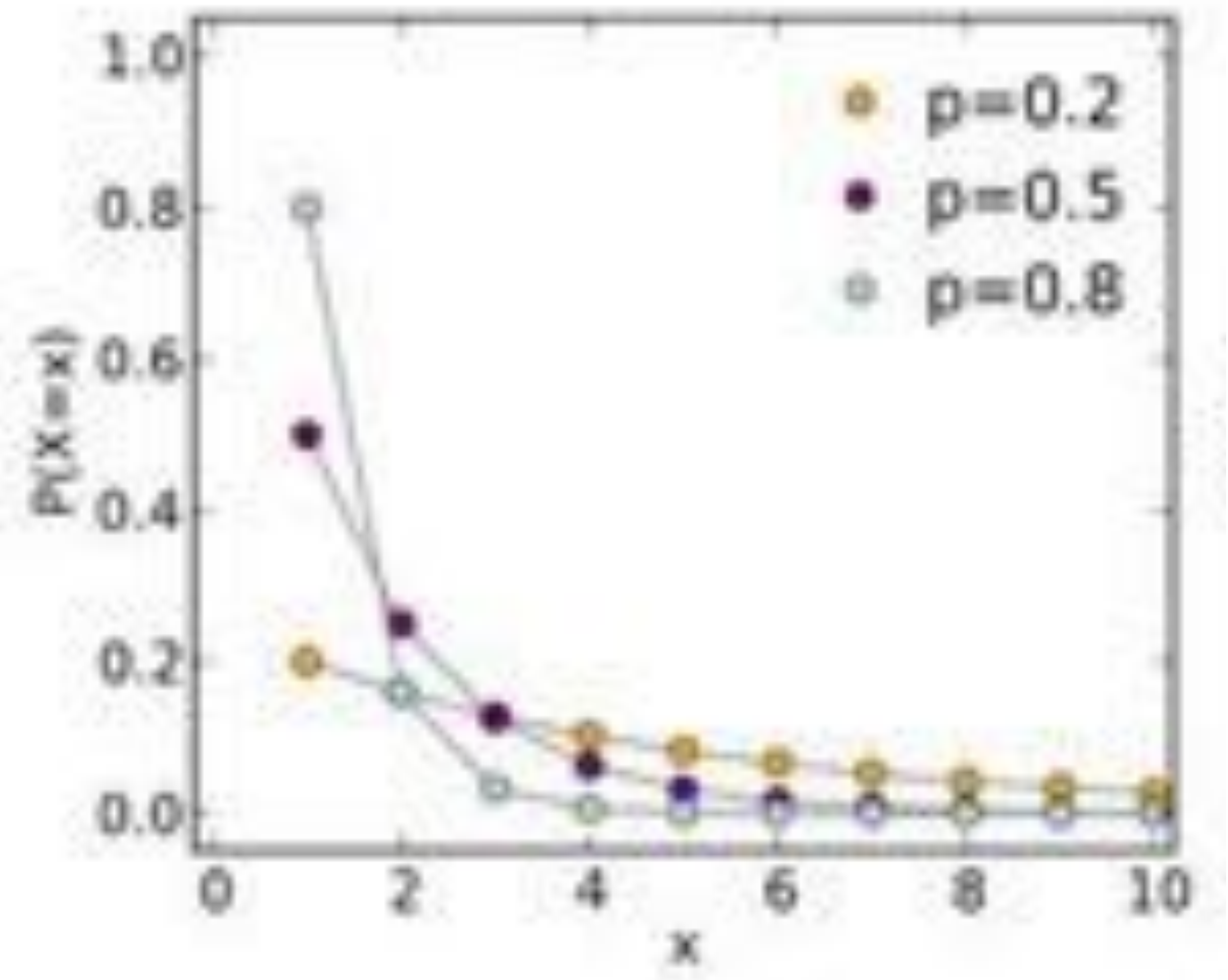
Flips	X
10101	1
01011	2
00011	4

n	$X_1,\ldots,X_n$	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	$X_1 = \ldots = X_{n-1} = 0 \quad X_n = 1$	$q^{n-1}p$

Geometric Distribution

G_p $0 < p \leq 1$

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



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

ΣWILL IT ADD?

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

$$\begin{aligned}(1 + q + q^2 + \ldots)(1 - q) &= 1 + \cancel{q} + \cancel{q^2} + \ldots \\ &\quad\; - \cancel{q} - \cancel{q^2} - \ldots\end{aligned}$$

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

YES IT ADDS!

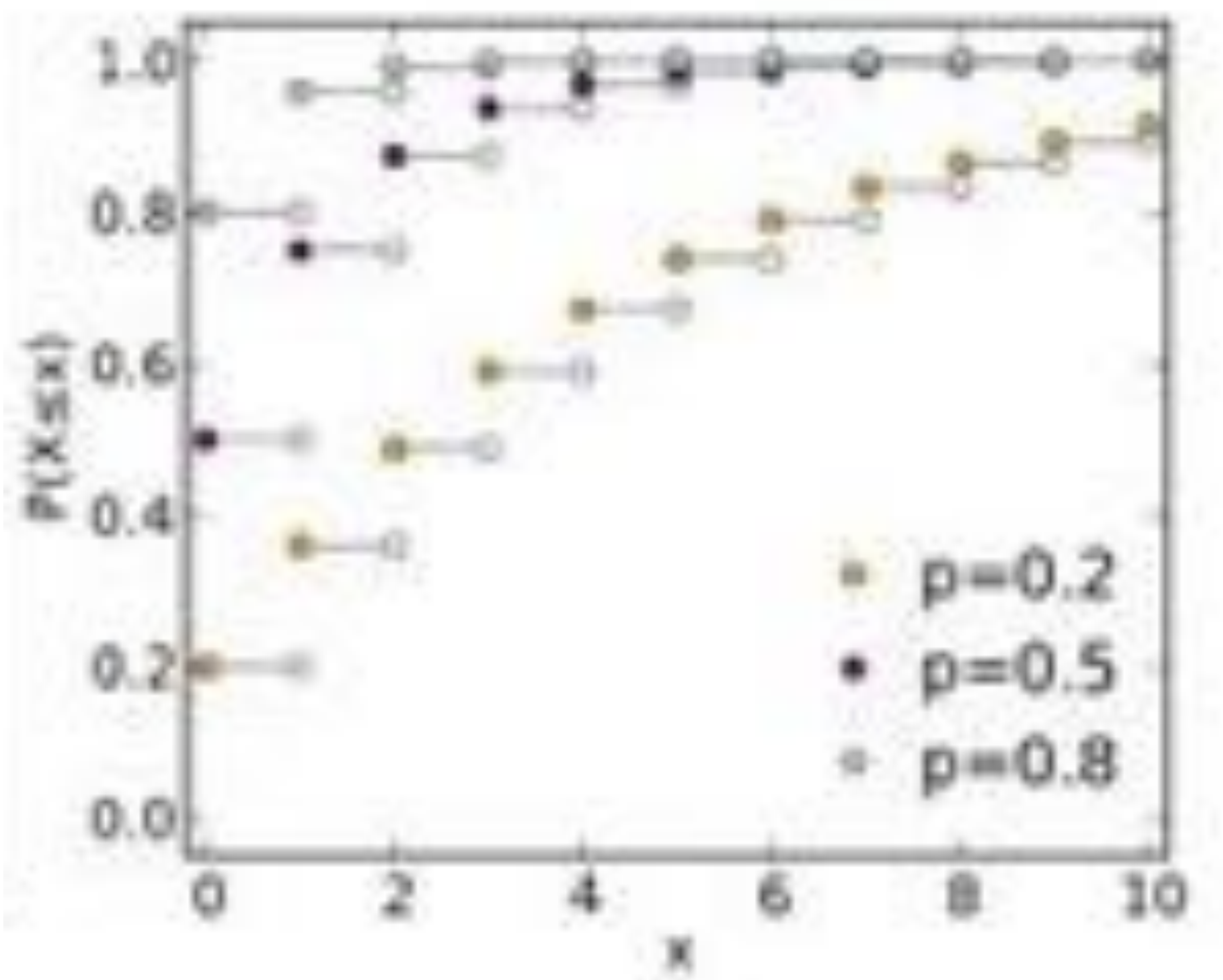
$$\sum_{n=1}^{\infty} p(1-p)^{n-1} = p \sum_{i=0}^{\infty} (1-p)^i = p \cdot \frac{1}{1-(1-p)} = \frac{p}{p} = 1$$

CDF

$n \in \mathbb{N}$

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

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



Expectation via "Right" CDF

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

$$\begin{array}{cccc}
P_1 & & & \\
P_2 & P_2 & & \\
P_3 & P_3 & P_3 & \\
P_4 & P_4 & P_4 & P_4 \\
\vdots & \vdots & \vdots & \vdots
\end{array}$$

$$E(X) = \sum_{k=0}^{\infty} k P_k$$

$$= P_1 + 2P_2 + 3P_3 + \ldots$$

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

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

$$EX(X-1) = \sum_{n=1}^{\infty} n(n-1) \cdot P(X=n) = p \sum_{n=2}^{\infty} n(n-1) q^{n-1}$$

$$= pq \sum_{n=2}^{\infty} \frac{d^2}{dq^2} q^n = pq \frac{d^2}{dq^2} \sum_{n=2}^{\infty} q^n$$

$$= pq \frac{d^2}{dq^2} \left(\frac{1}{1-q} - 1 - q \right)$$

$$= pq \frac{2}{(1-q)^3} = \frac{2q}{p^2}$$

$$\left(\frac{1}{1-q}\right)' = \frac{1}{(1-q)^2}$$

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

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

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

Fair Coin

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

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

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

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

Memoryless

A distribution over $\mathbb{P} = \{1,2,..\}$ 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 $p \overset{\text{def}}{=} P(X = 1), \qquad q \overset{\text{def}}{=} 1 - p = P(X > 1)$

$$\forall\, n \geq 1, \quad P(X = n+1) = P(X > 1 \ \wedge \ X = n+1)$$

$$= P(X > 1) \cdot P(X = n+1 | X > 1)$$

$$= q \cdot P(X = n)$$

Hence P(X=2)=qp, P(X=3)=q^2p, ….

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

r Successes

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

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

$n \geq r$

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

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

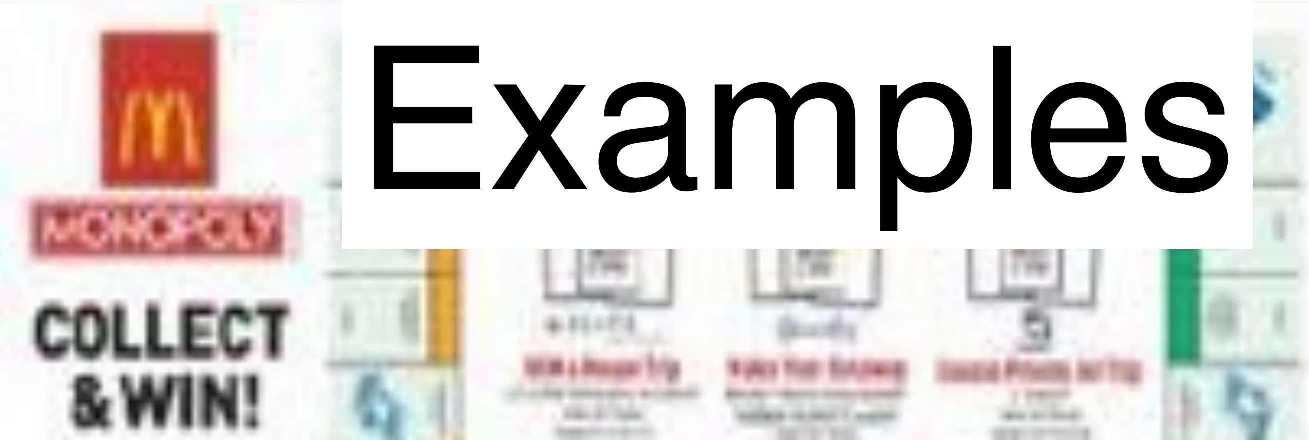
Memoryless

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

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

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

Next: Examples



Geometric Distribution Examples

Startup Statistics

P(startup success) = 20%, independent of previous attempts

Expected # startups till first success

$X \sim G_{0.2}$ $\qquad E(X) = \frac{1}{.2} = 5$

Home-Grown Entrepreneur

One of first three startups succeeds

Dad will fund up to three startups P(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)?

$$P(X \in \{4,5,6\}|X > 3) = P(4|X > 3) + P(5|X > 3) + P(6|X > 3)$$

$$= P(1) + P(2) + P(3) = P(X \leq 3) \approx 49\%$$

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

$$P(3 < X \leq 6) = P(X > 3 \cap X \leq 6) = P(X > 3) \cdot P(x \leq 6|x > 3)$$

$$= (0.8)^3 \cdot 0.49 \approx 25\%$$

$P(X_1, X_2, X_3 \text{ failed}) = q^3$

$$P(3 < X \leq 6) = F(6) - F(3) = (1 - 0.8^6) - (1 - 0.8^3)$$

$$= 0.8^3 - 0.8^6 \approx 25\%$$

Foreign-Born Entrepreneur

X - time to first success   p=0.2

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

$$E(\, r^X\,) = \sum_{k=1}^{\infty} r^k P(X=k) = \sum_{k=1}^{\infty} pq^{k-1} r^k = pr \sum_{i=0}^{\infty} (qr)^i$$

$$= \frac{pr}{1-qr} = \frac{0.2 \cdot 0.5}{1-0.8 \cdot 0.5} = \frac{0.1}{1-0.4} = \frac{0.1}{0.6} \approx 16.67\%$$

Coupon Collector Problem

Pre GROUPON

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?

Coupon $\underbrace{2}\ \underbrace{2\ \ 3}\ \underbrace{2\ \ 3\ \ 3\ \ 1}$

$X_1 = 1$ $X_2 = 3 - 1 = 2$ $X_3 = 7 - 3 = 4$

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

$$X_i \sim G(\tfrac{n-(i-1)}{n}) = G(\tfrac{n-i+1}{n})$$

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

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

$$EX = \sum_{i=1}^{n} EX_i = \sum_{i=1}^{n} \tfrac{n}{n-i+1} = \tfrac{n}{n} + \tfrac{n}{n-1} + \tfrac{n}{n-2} + \ldots + \tfrac{n}{1}$$

$$= n(\tfrac{1}{1} + \tfrac{1}{2} + \ldots + \tfrac{1}{n}) = nH_n \approx n \ln n + 0.577n$$

Harmonic Sum $H_n = \tfrac{1}{1} + \tfrac{1}{2} + \ldots + \tfrac{1}{n} \to \ln n + 0.577\ldots$

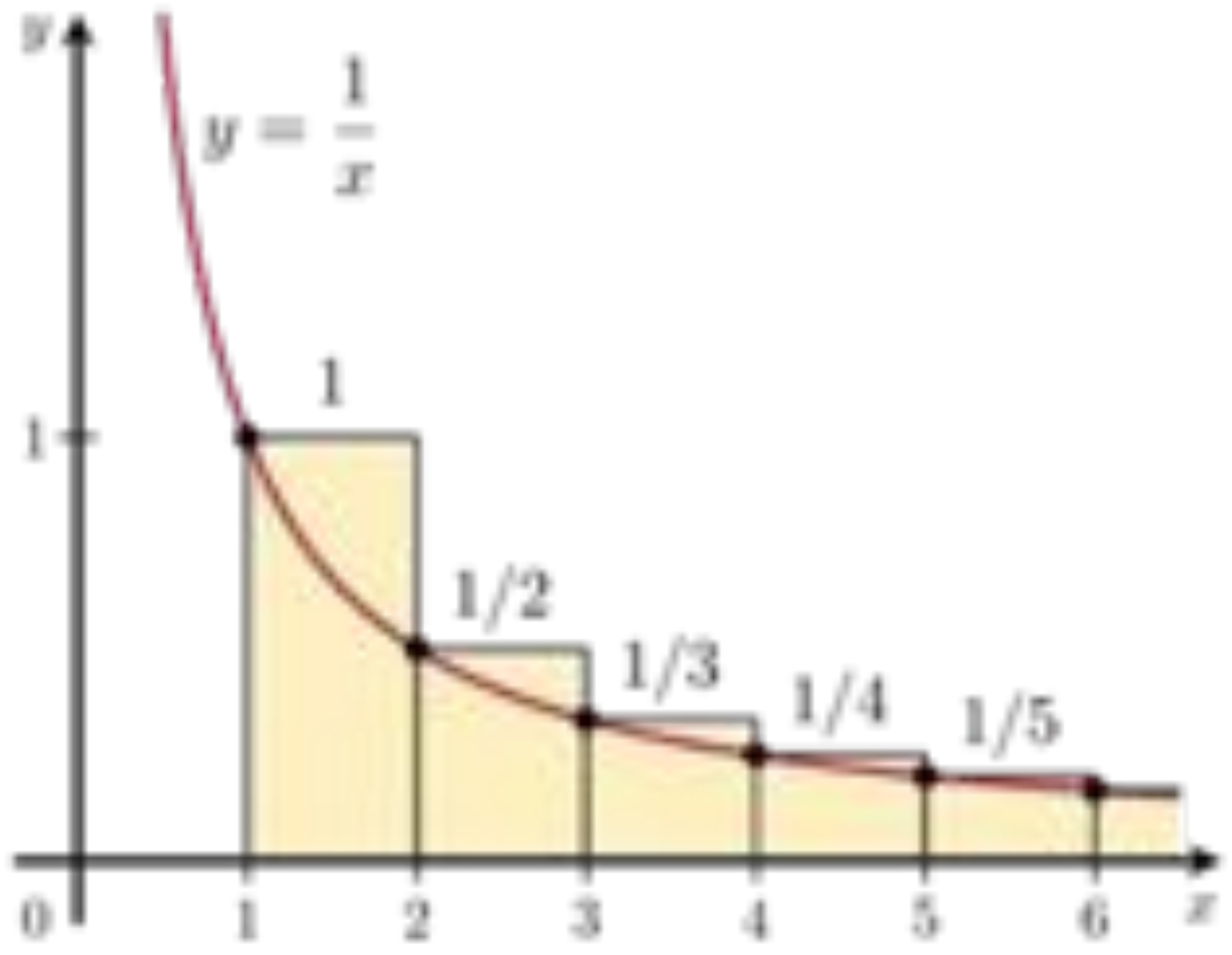
Harmonic Sum

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

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

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

$$H_n \to \ln n + 0.577\ldots$$



Variance

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

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

$$\overset{\perp\!\!\!\perp}{=} \sum_{i=1}^{n} V(X_i)$$

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

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

$$\leq \frac{\pi^2}{6} n^2 \qquad\qquad \sigma \leq \frac{\pi}{\sqrt{6}} n$$

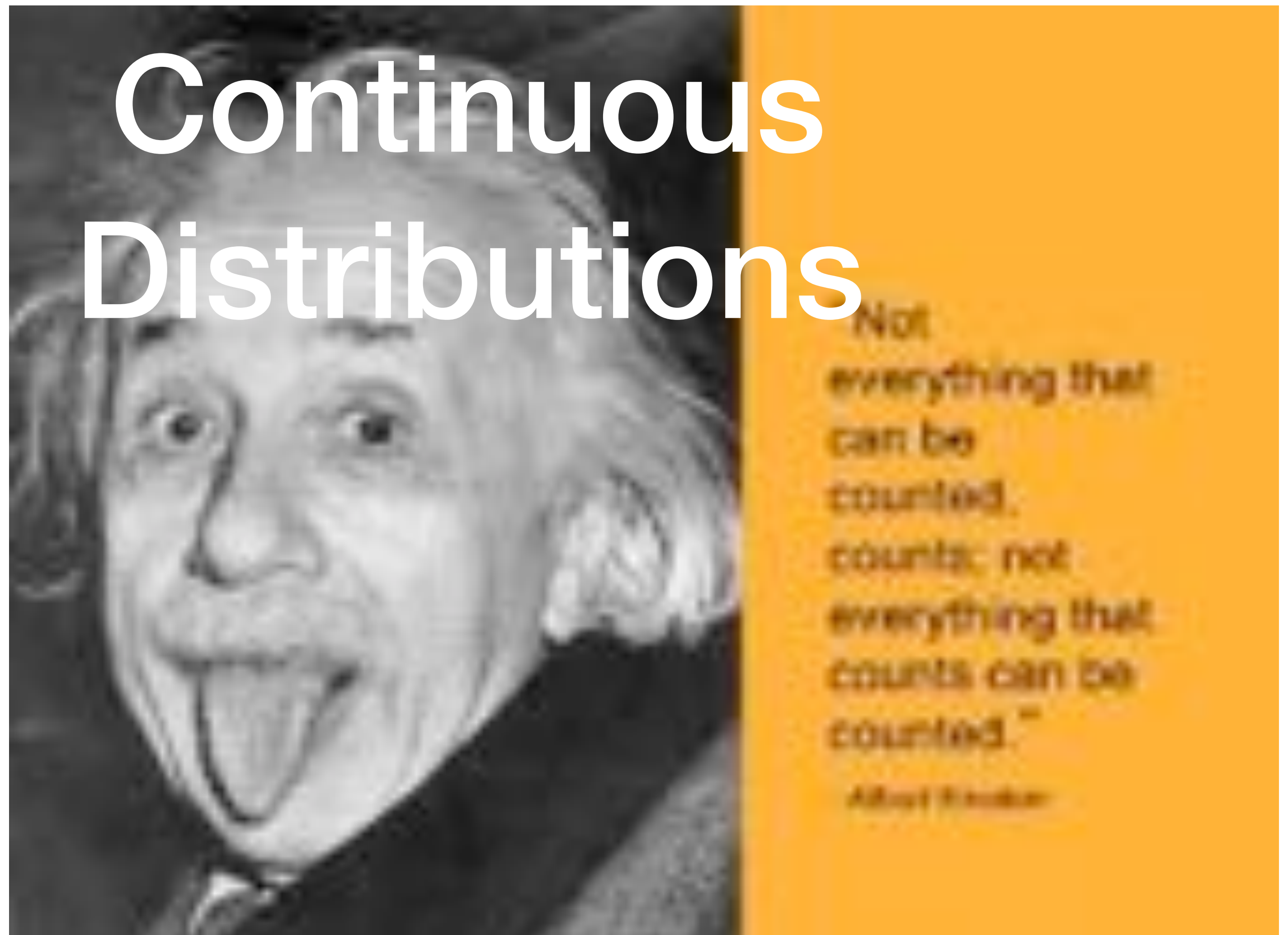
Summary

Geometric-distribution examples

Coupon collector problem

Discrete distribution families

Bernoulli, Binomial, Poisson, Geometric

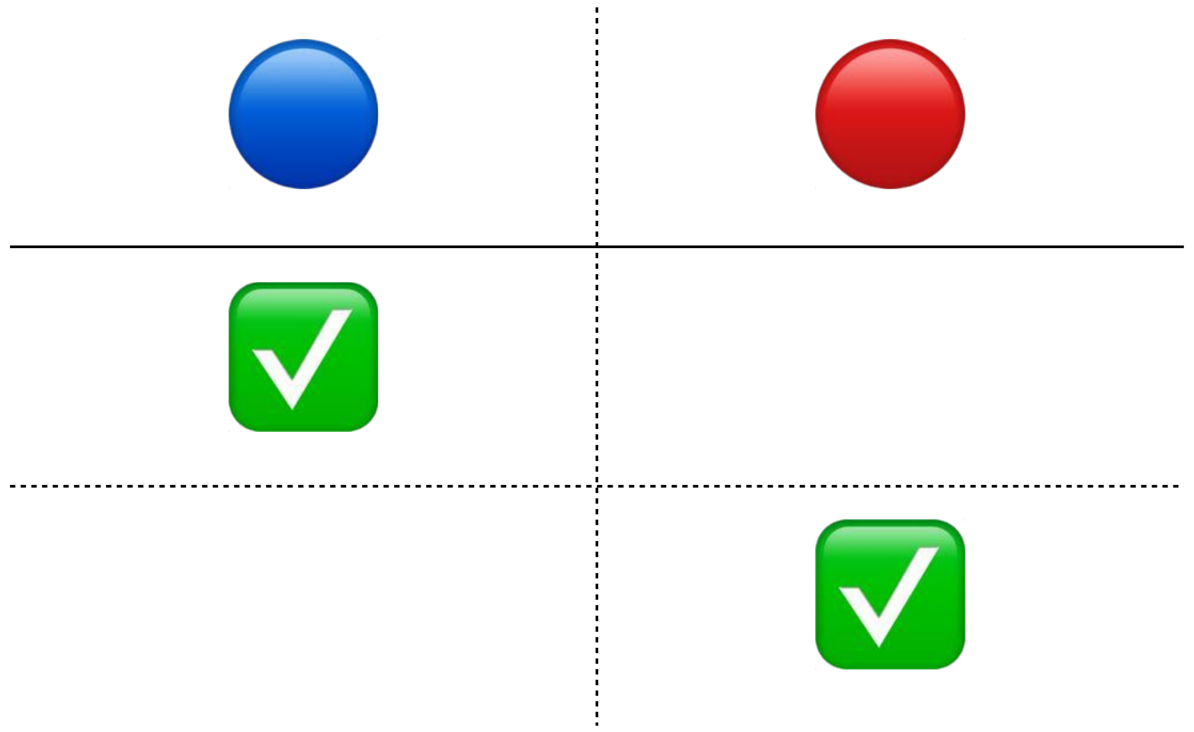


Continuous Distributions

Hypergeometric Distribution

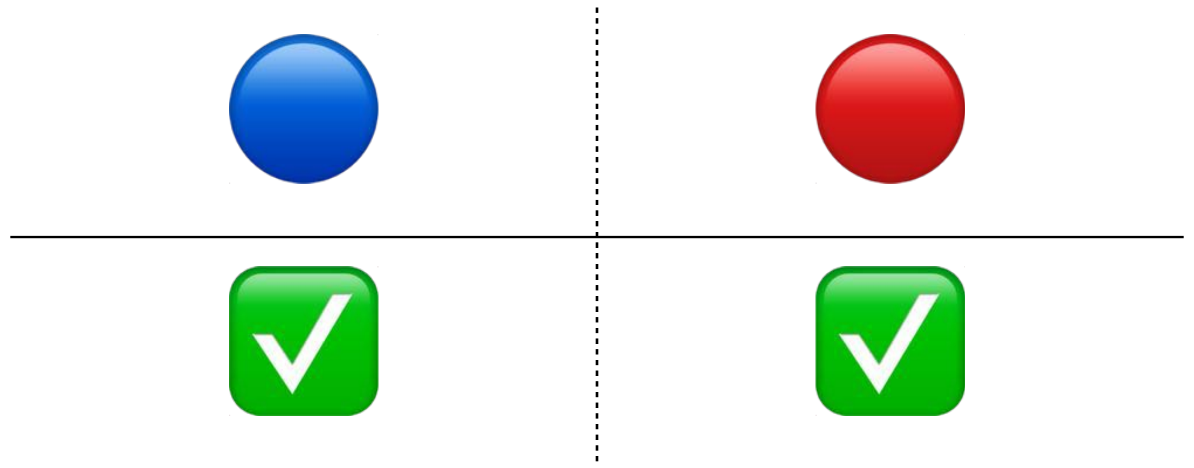
b = r = 1 n = 1

k	P(k)
0	½
1	½



b = r = 1 n = 2

k	P(k)
0	0
1	1
2	0



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

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

$$P(2) = 0$$

b = 1 r = 2 n = 2

k	P(k)
0	⅓
1	⅔
2	0

🔵	🔴	🔴
✅	✅	
✅		✅
	✅	✅

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

Hypergeometric Distribution

b blue balls   r red balls

Pick n ≥ 0 balls

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

If k blue balls, then n-k red balls

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

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 $\rightarrow$ $\dfrac{\binom{3}{3}\binom{4}{2}}{21} = \dfrac{6}{21}$

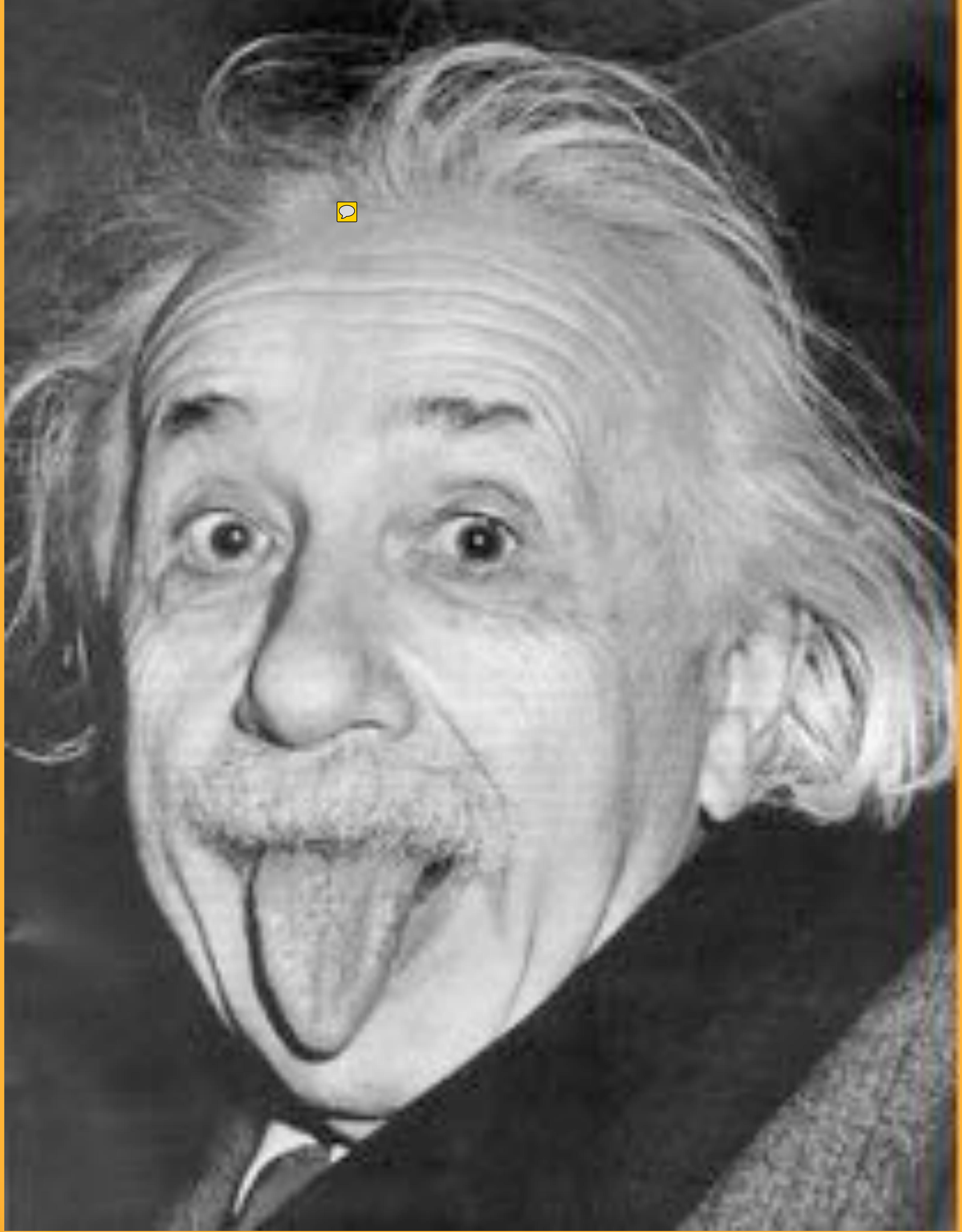
3 white $\rightarrow$ 2 black $\rightarrow$ $\dfrac{\binom{3}{2}\binom{4}{3}}{21} = \dfrac{12}{21}$

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

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

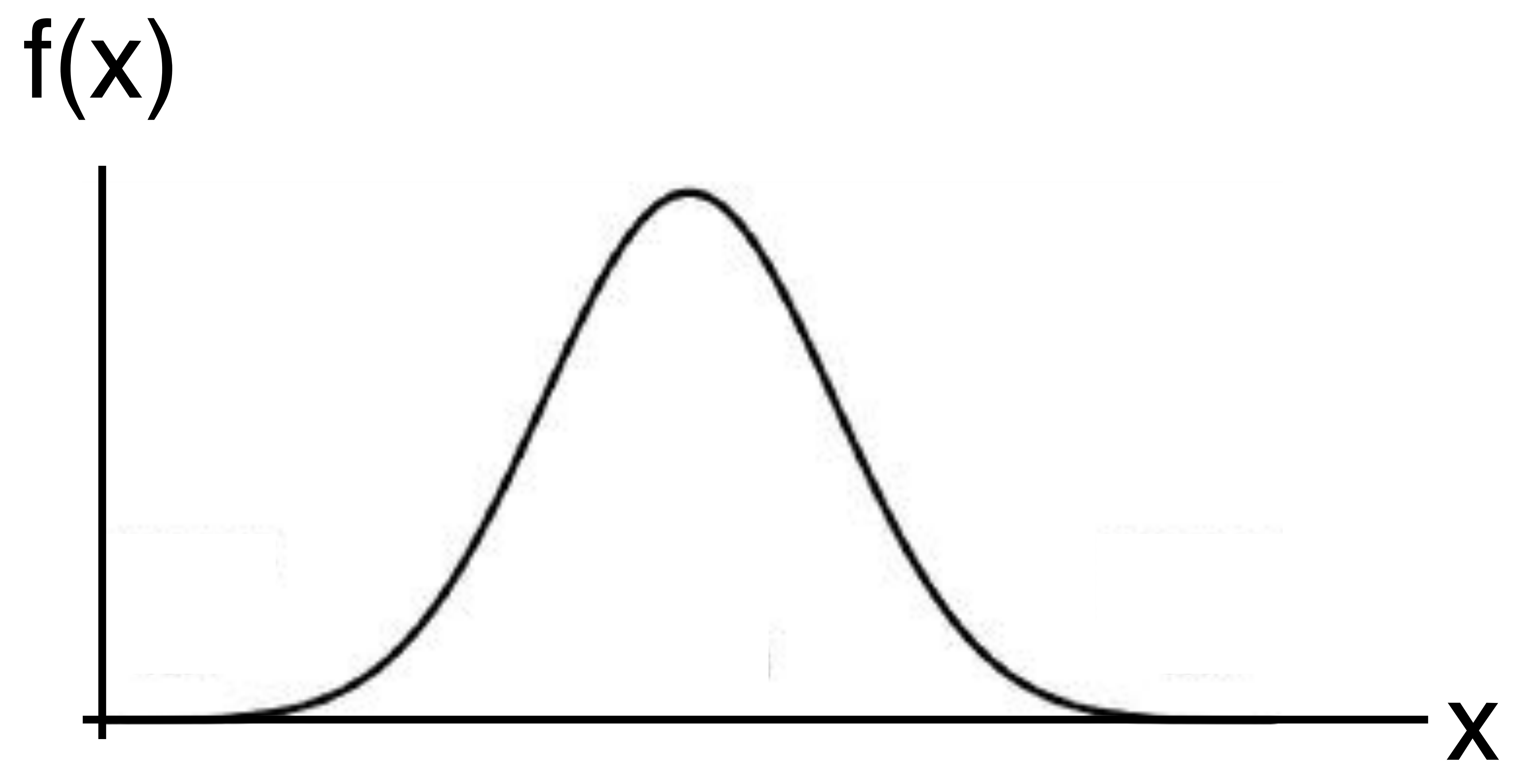
- **Cost**: stock, house, pork bellies
- **Rates**: interest, exchange, unemployment

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)

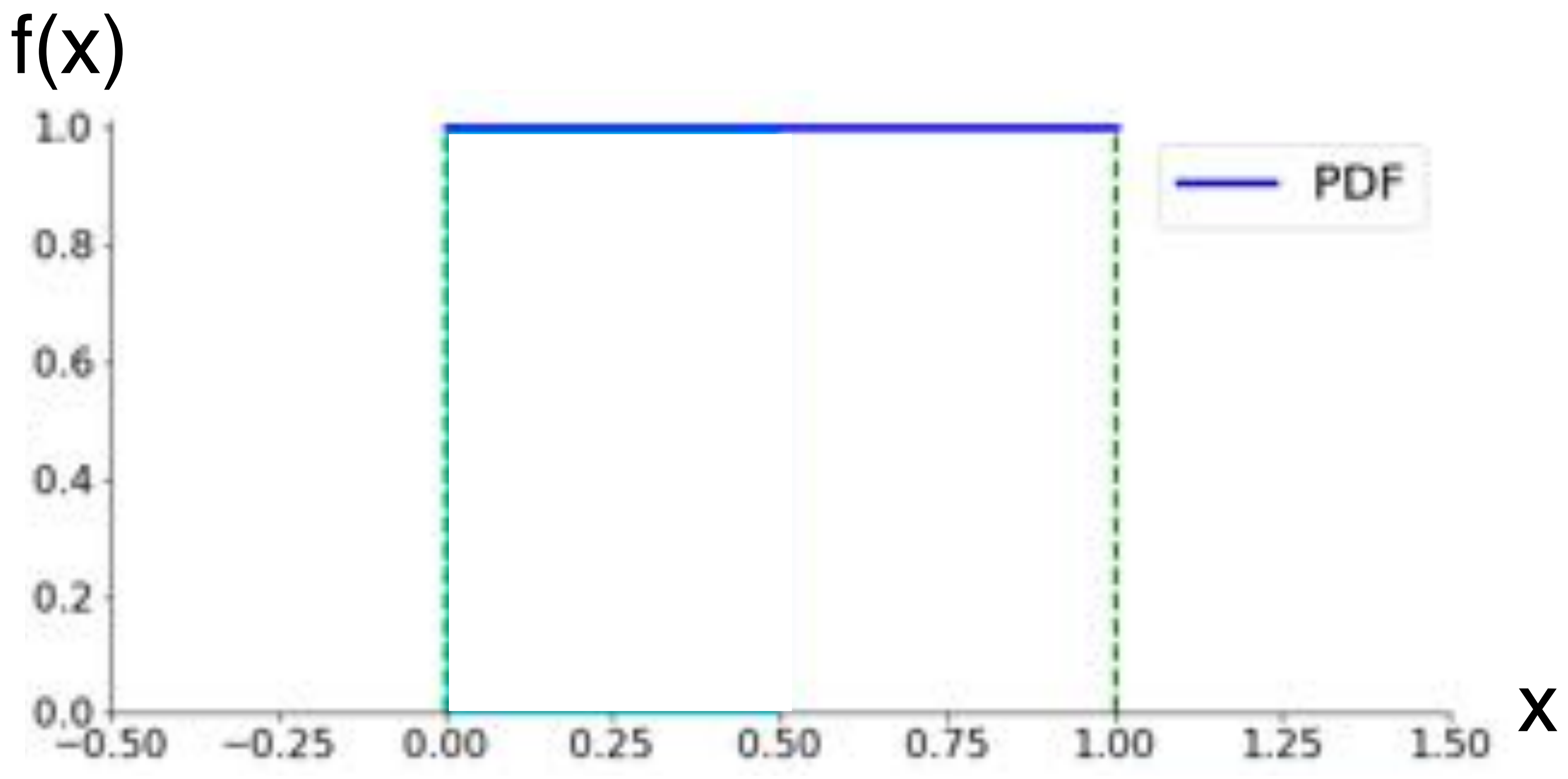


Comparison to Discrete

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

Uniform

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

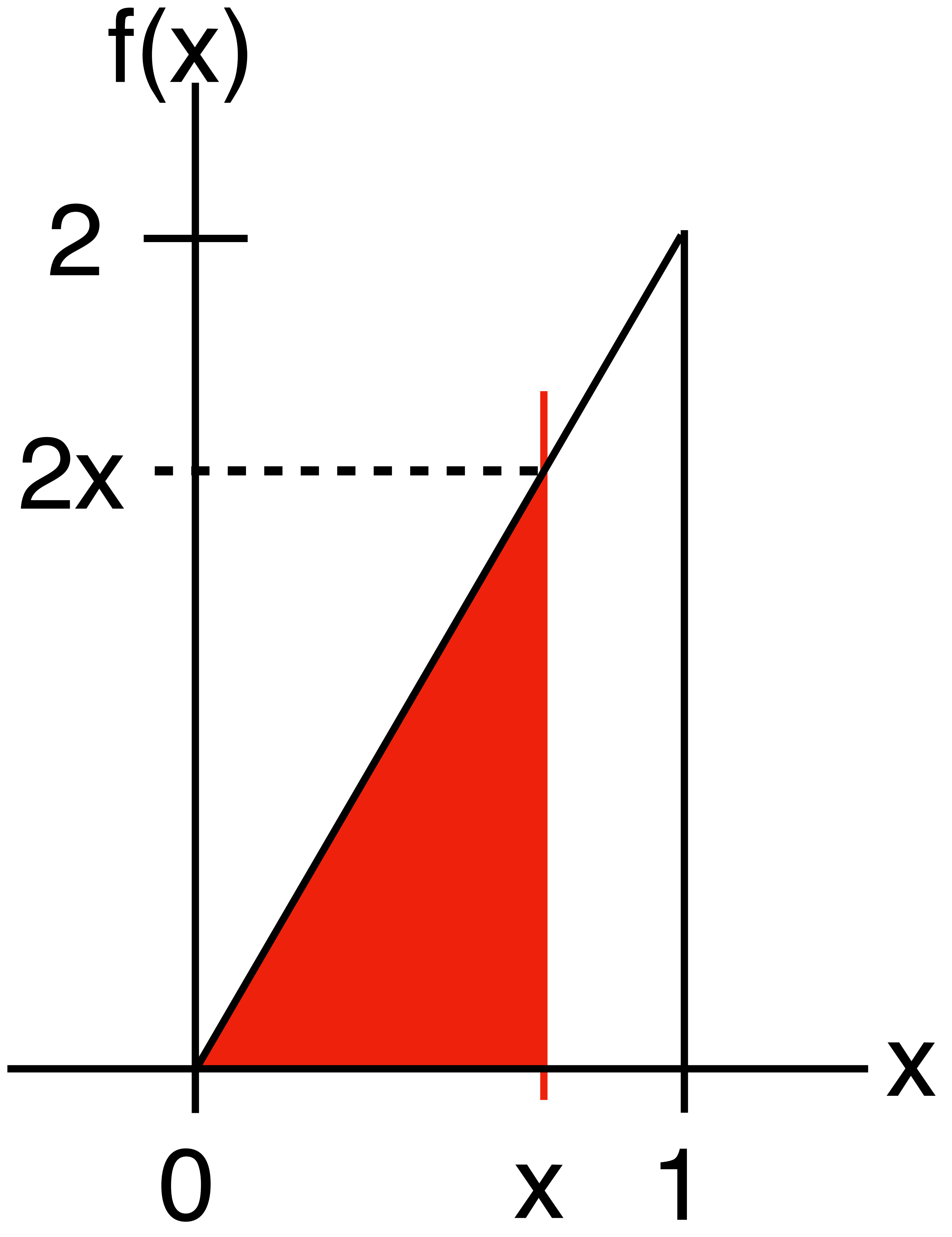


Will it $\int$?   Area = 1 · 1 = 1

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

Triangle

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



Will it $\int$?

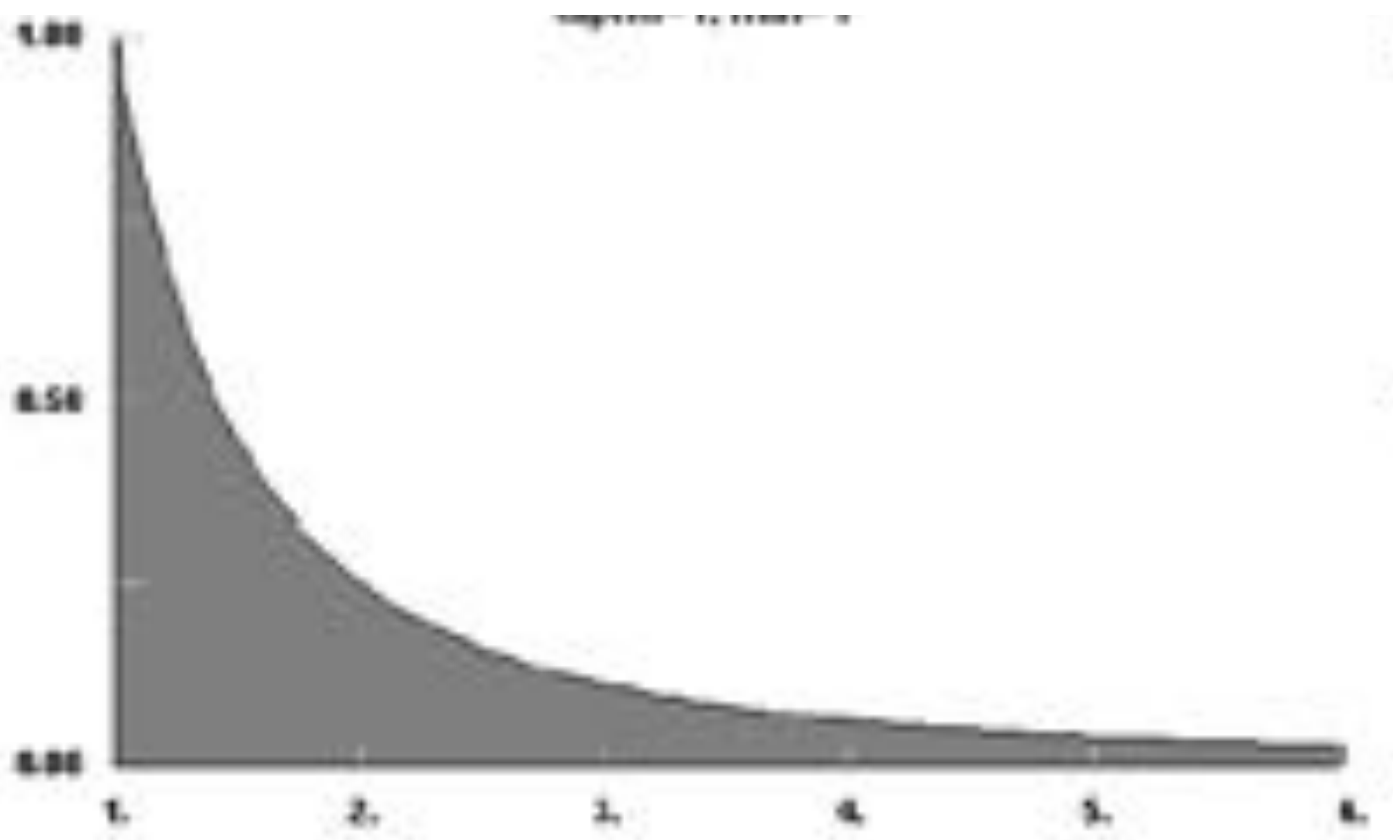
Area $= 2 \cdot 1 \cdot \frac{1}{2} = 1$

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

Infinite Support

Power law

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



Will it ∑?

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

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 ≤ X ≤ b)

Area between a and b

P(X ≤ b) - P(X ≤ a)

Cumulative distribution function

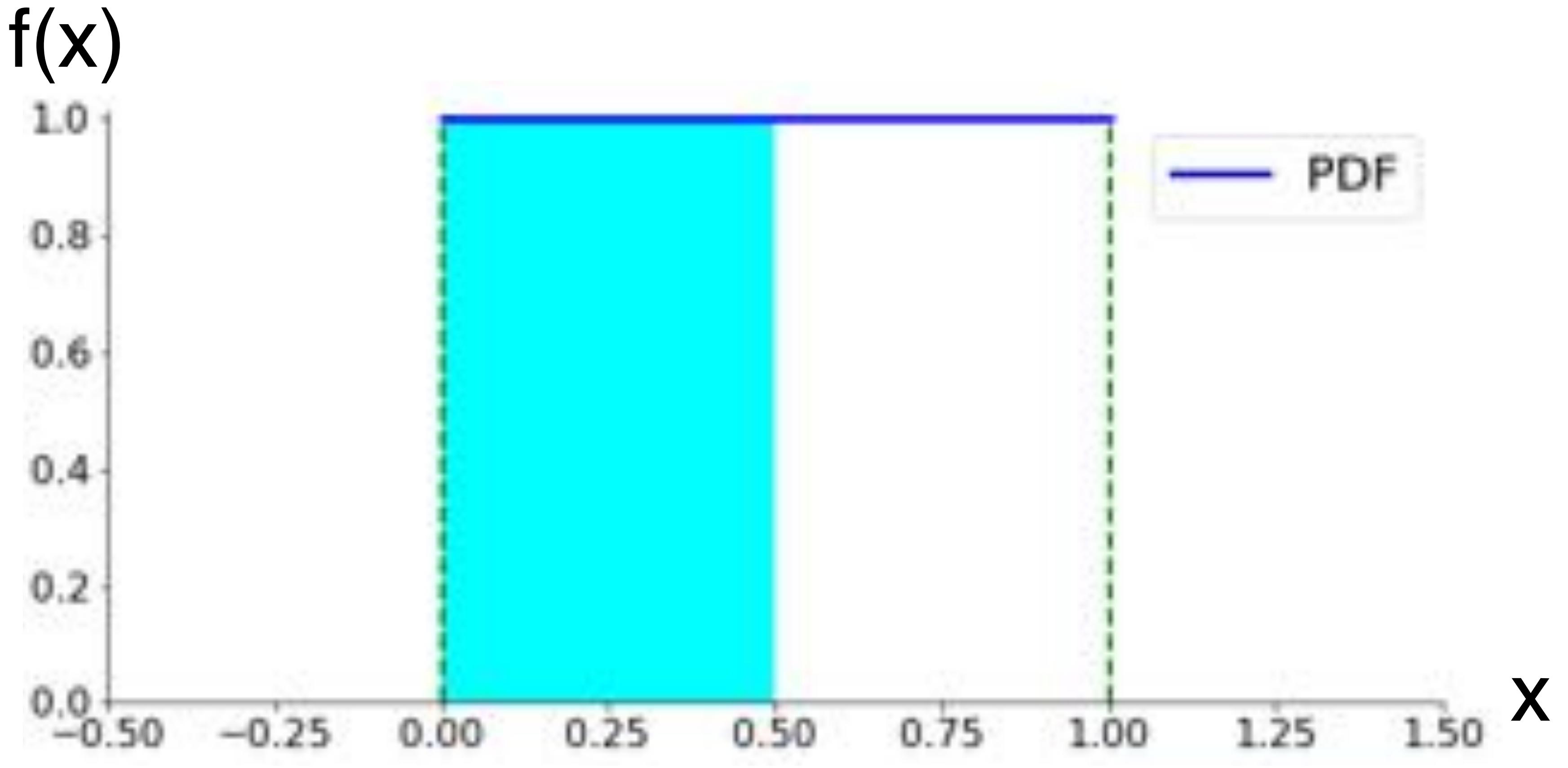
Cumulative Distribution Function (CDF)

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

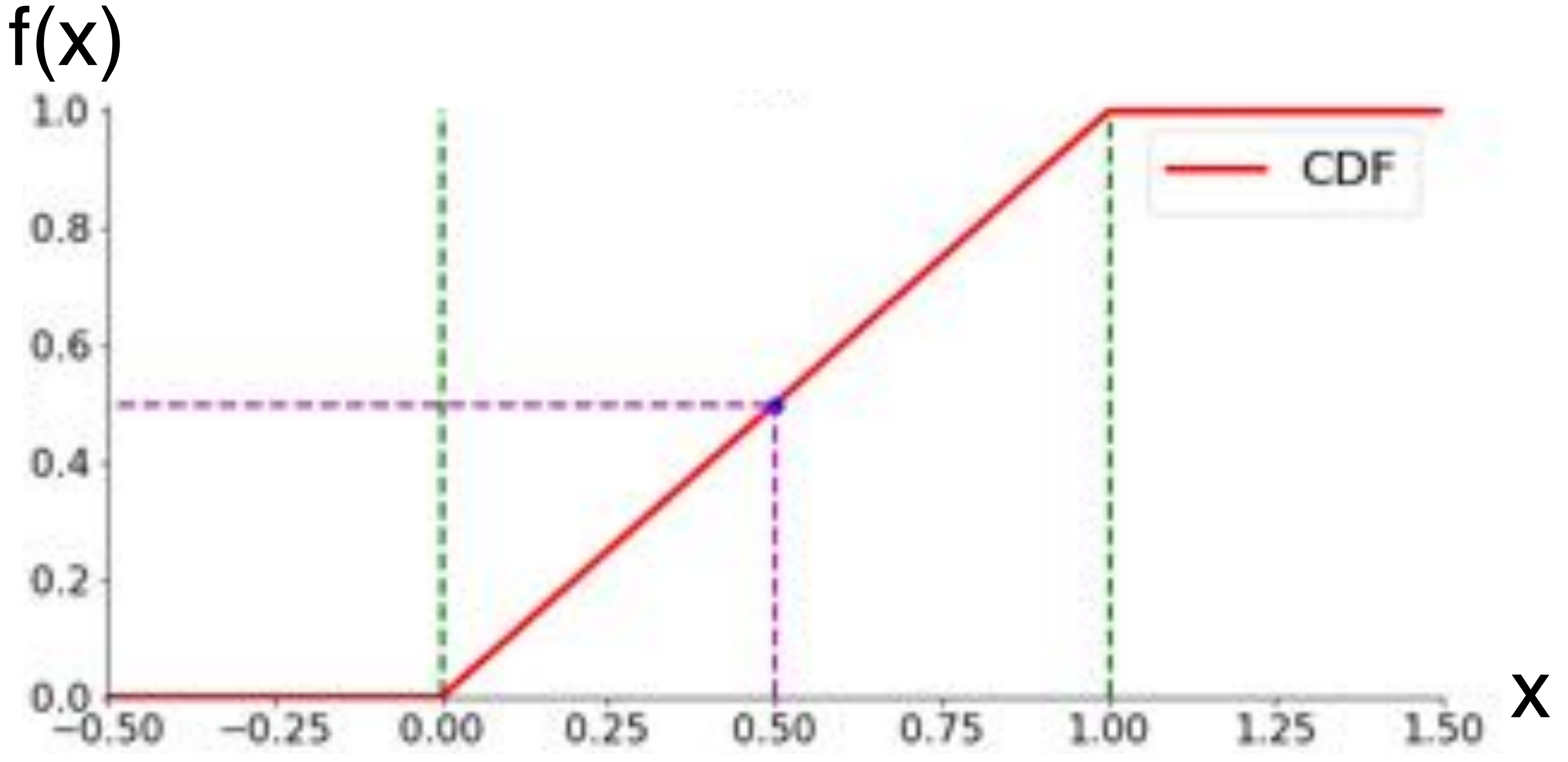
	Discrete	Continuous
PF → CDF	$\sum_{u \leq x} p(u)$	$\int_{-\infty}^{x} f(u)du$
CDF → PF	$p(x) = F(x) - F(x^*)$	$f(x) = F'(x)$

x^* - element preceding x

Uniform



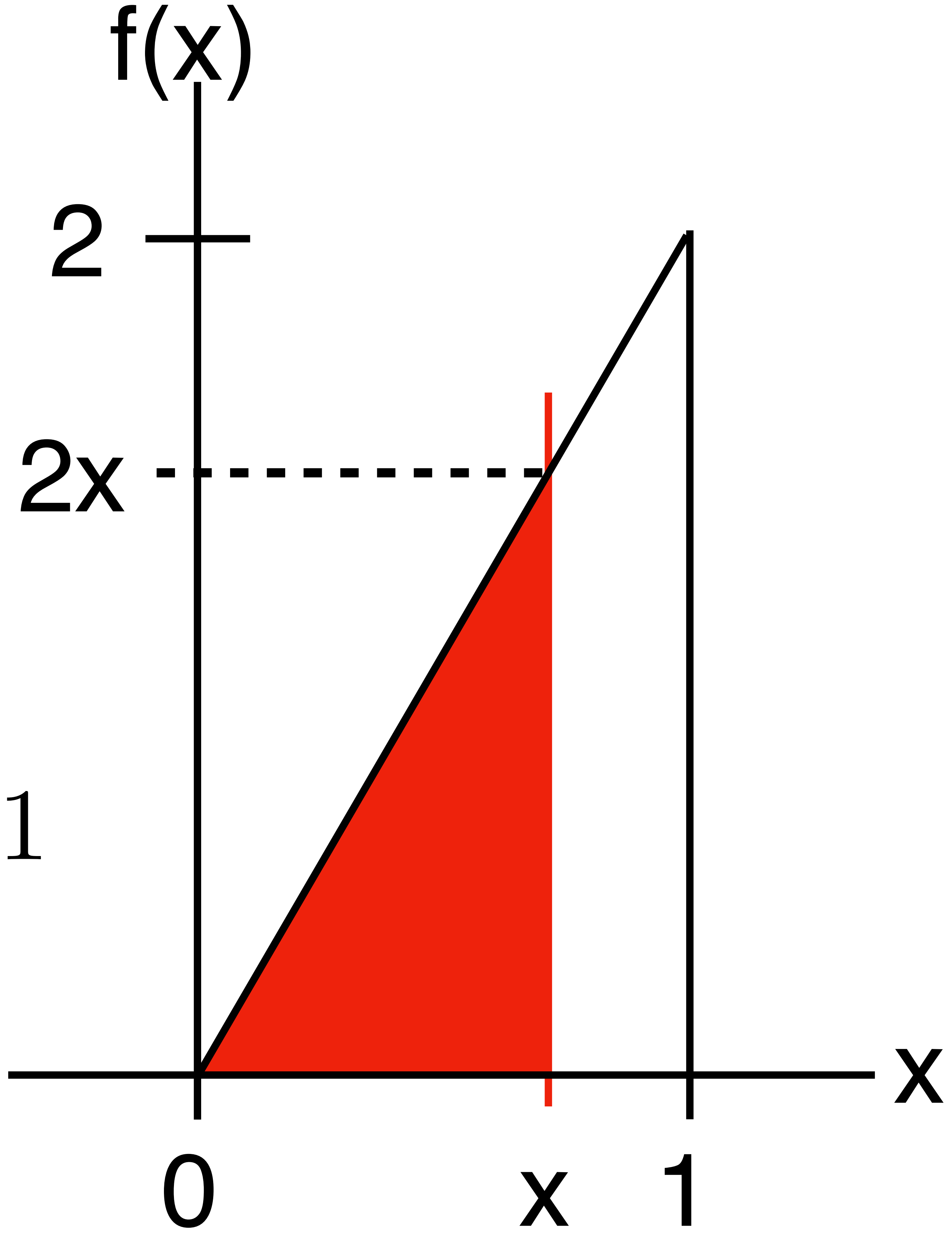
$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} 0 & x \leq 0 \\ \int_0^x 1du = u\Big|_0^x = x & 0 \leq x \leq 1 \\ 1 & 1 \leq x \end{cases}$$



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

Triangle

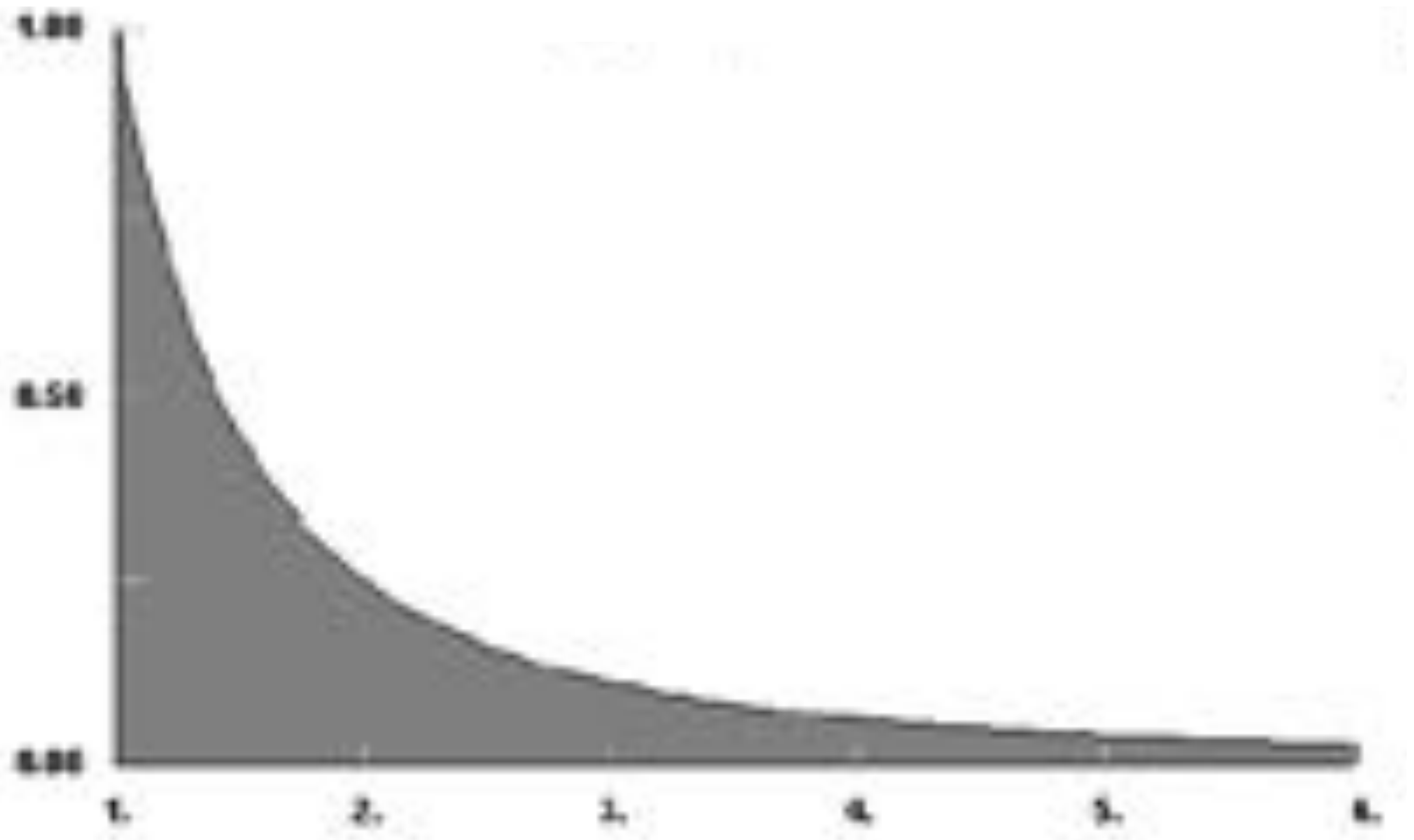
$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} 0 & x \leq 0 \\ \int_0^x 2udu = u^2\Big|_0^x = x^2 & 0 \leq x \leq 1 \\ 1 & 1 \leq x \end{cases}$$



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

Infinite Support

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



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

Properties of the CDF

F(x) = integral

Nondecreasing

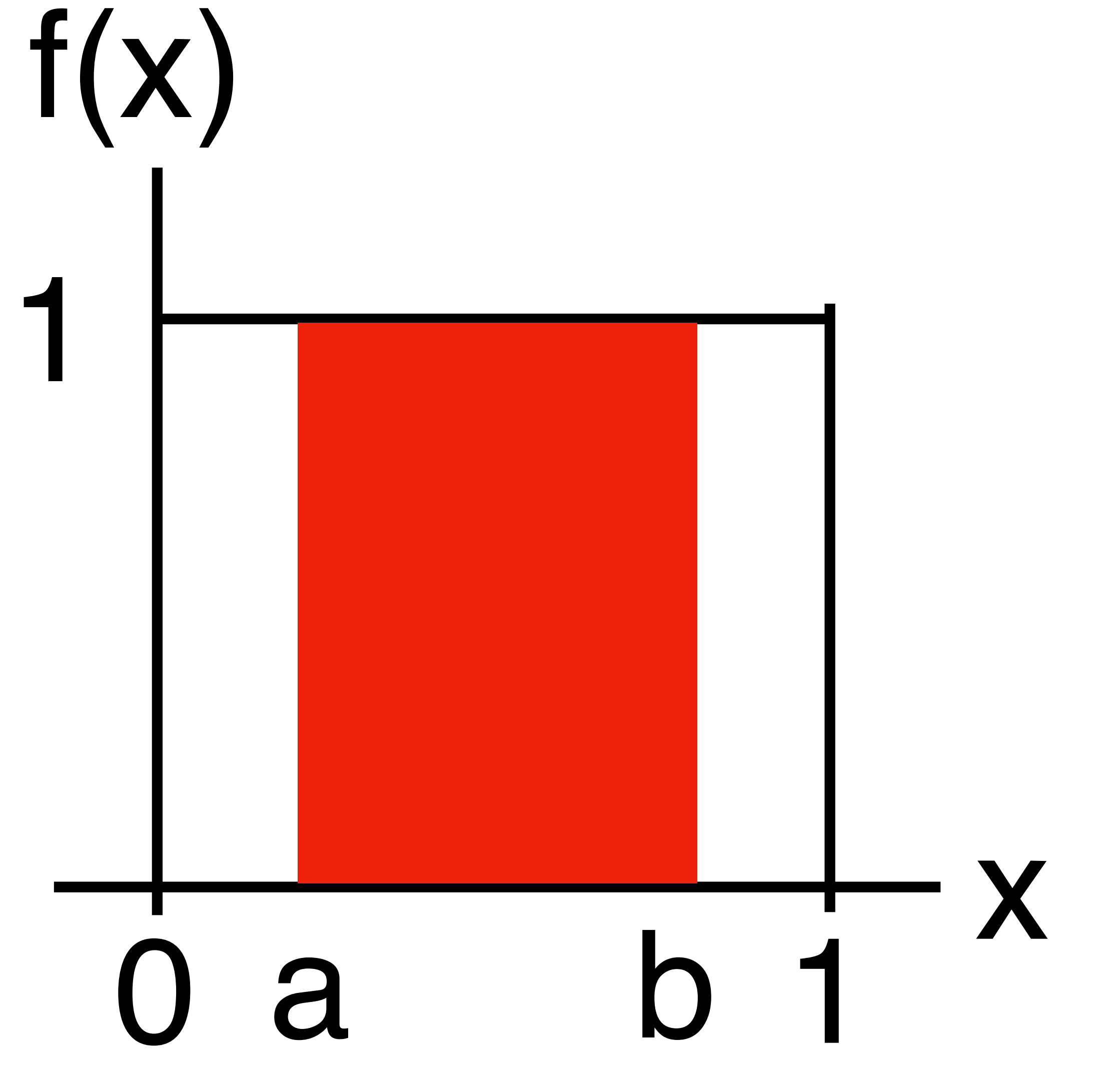
$F(-\infty)=0$

$F(\infty)=1$

Continuous

Interval Probabilities

Examples



Uniform

$0 \le a \le b \le 1$

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

$$P(0.6 \le X \le 1.3) = P(0.6 \le X \le 1) = 0.4$$

$$= F(1.3) - F(0.6) = 1 - 0.6 = 0.4$$

Power law

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

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

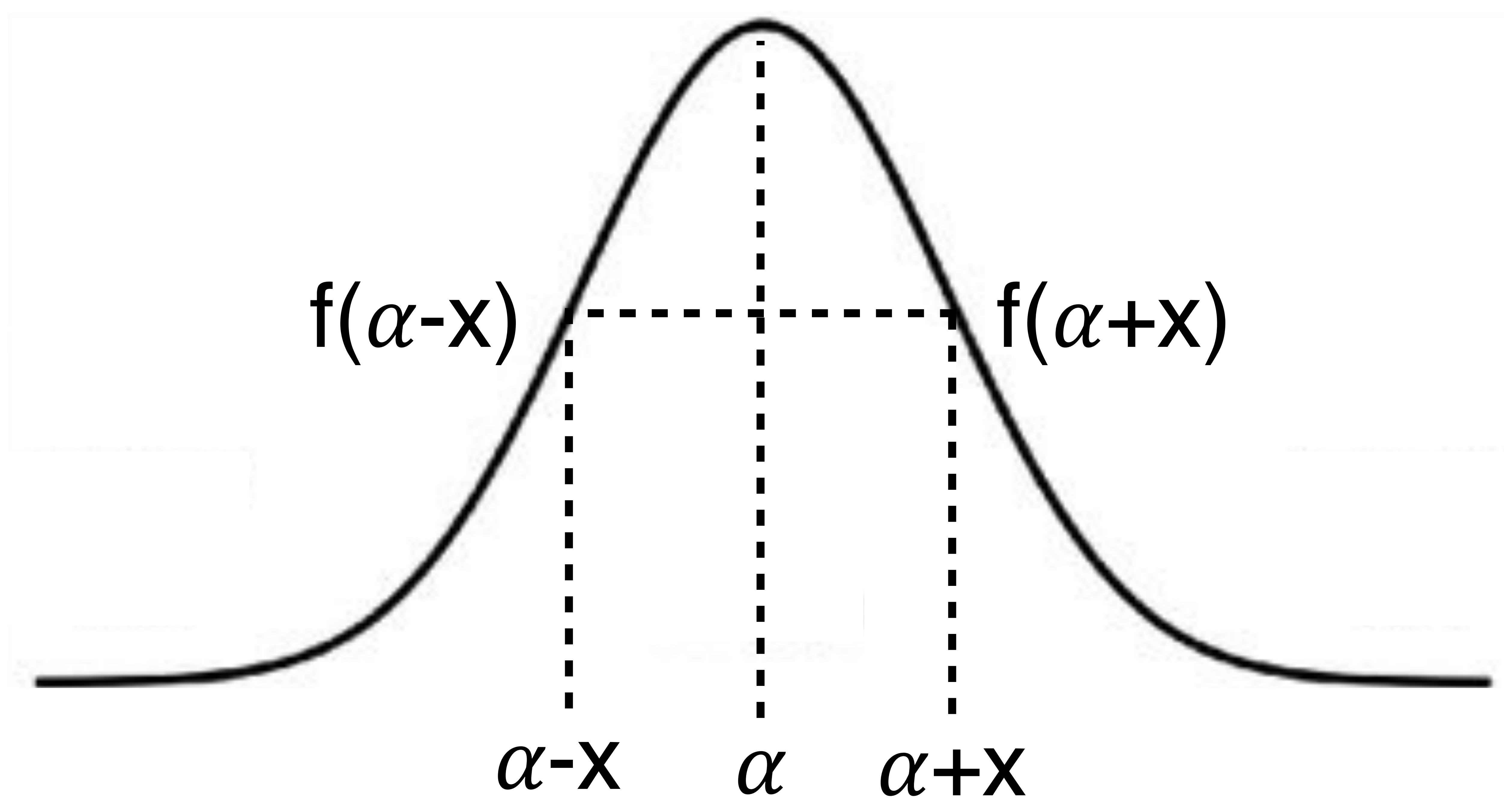
	Discrete	Continuous
EX	$\sum x \cdot p(x)$	$\int_{-\infty}^{\infty} x f(x) dx$

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

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

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

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

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

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

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

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

Discrete vs. Continuous

	Discrete	Continuous
Prob. Fun.	pmf - p	pdf - f
≥ 0	$p(x) \geq 0$	$f(x) \geq 0$
∑ = 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)	$\sum_{u \leq x} p(u)$	$\int_{-\infty}^{x} f(u)du$
μ = 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) ≤ 1 f can be larger

f(x) = F'(x)

P(a ≤ X ≤ b) = P(a < X < b) = F(b) - F(a)

Functions of Continuous Random Variables