

CS 215

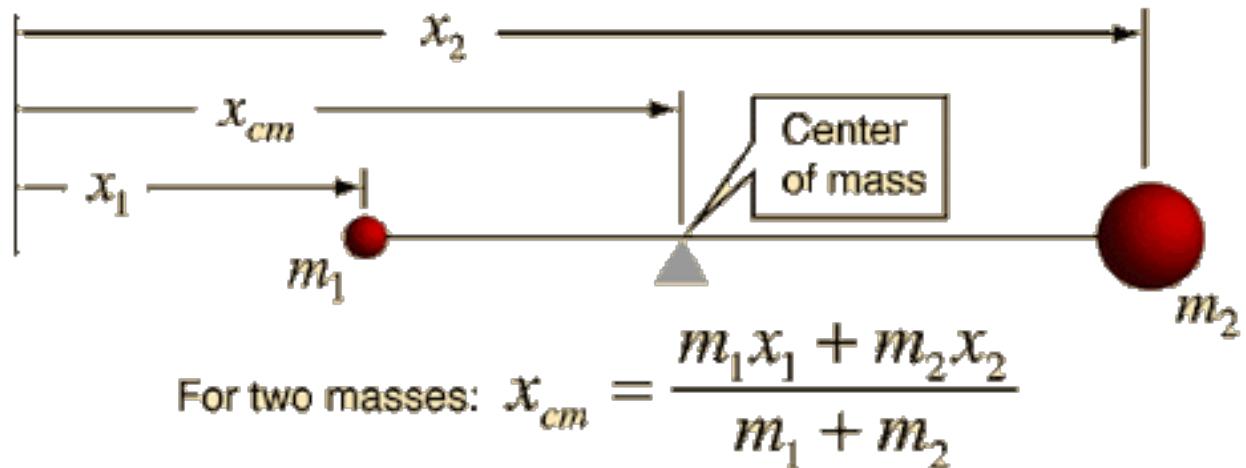
Data Analysis and Interpretation

Expectation

Suyash P. Awate

Expectation

- “Expectation” of the random variable;
“Expected value” of the random variable;
“Mean” of the random variable.
- “Expected value” isn’t the value that is most likely to be observed in the random experiment
- Can think of it as the center of mass of the probability mass/density function



Expectation

- Definition:

Expectation of a Discrete Random Variable: $E[X] := \sum_i x_i P(X = x_i)$

- Frequentist interpretation of probabilities and expectation
 - If a random experiment is repeated infinitely many times, then the proportion of number of times event E occurs is the **probability** $P(E)$
 - If a random experiment underlying a discrete random variable X is repeated infinitely many times, then the proportion of number of experiments when X takes value x is $P(X=x)$
 - So, in $N \rightarrow \infty$ experiments, number of times X takes value x_i will $\rightarrow N.P(X=x_i)$
 - So, across all $N \rightarrow \infty$ experiments, **arithmetic average of observed values** will
 $\rightarrow (1/N) \sum_i (x_i) (N.P(X=x_i))$
 $= E[X]$

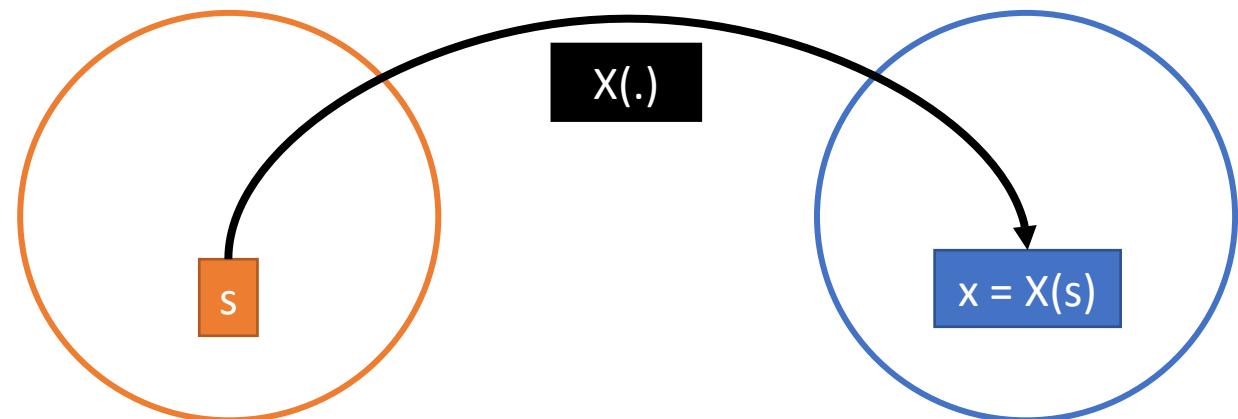
Expectation

- Another Formulation of Expectation

- Recall:

- **Discrete** random variable X is a function defined on a probability space $\{\Omega, \mathcal{B}, P\}$
- Function $X: \Omega \rightarrow \mathbb{R}$, maps each element in sample space Ω to a single numerical value belonging to the set of real numbers

$$\begin{aligned} E[X] &:= \sum_i x_i P(X = x_i) \\ &= \sum_i x_i \left(\sum_{s \in \Omega: X(s) = x_i} P(s) \right) \\ &= \sum_{s \in \Omega} X(s) P(s) \end{aligned}$$



Expectation

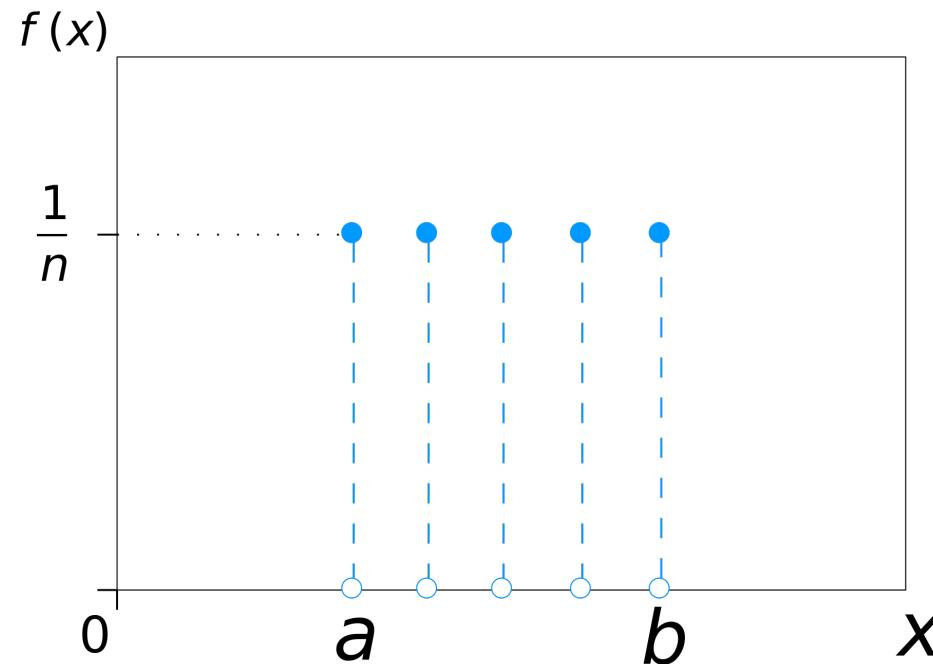
- Example

- “Expected value” for the uniform random variable modelling die roll

- Values on die are $\{1, 2, 3, 4, 5, 6\}$
- $E[X] = 3.5$

- **Expectation of a uniform random variable (discrete case)**

- If X has uniform distribution over n consecutive integers over $[a, b]$,
then $E[X] = (a+b)/2$

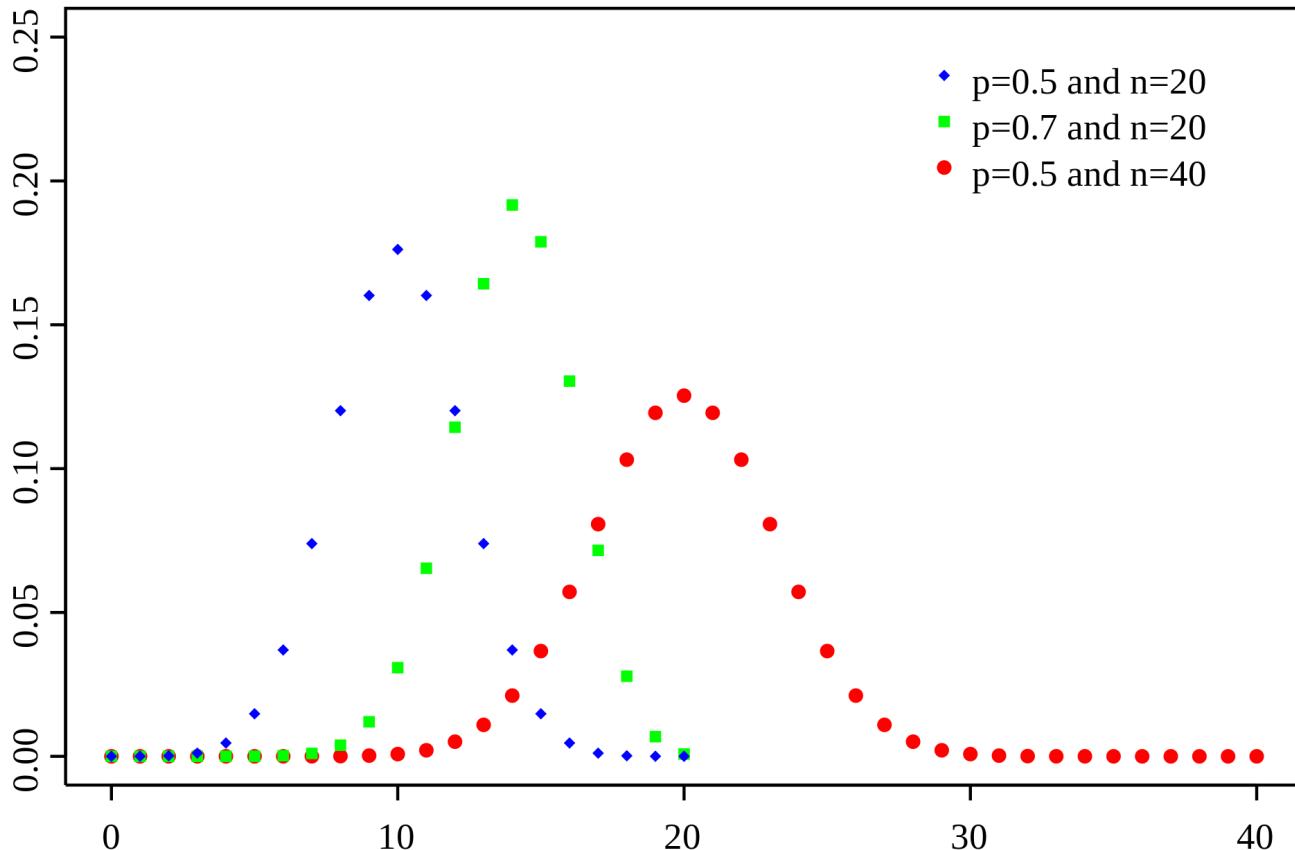


Expectation

- Example

- **Expectation of a binomial random variable (when n=1, this is Bernoulli)**

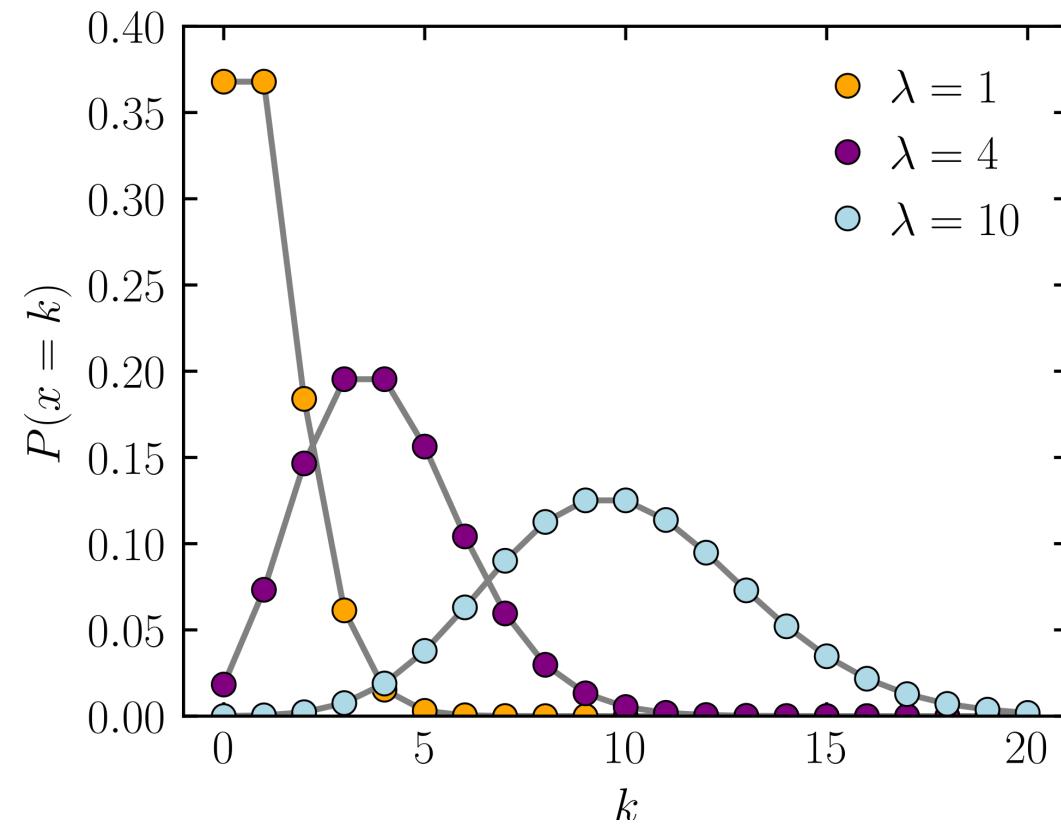
$$\begin{aligned} E(X) &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=1}^n k \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=1}^n n \binom{n-1}{k-1} p^k q^{n-k} \\ &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} \\ &= np \sum_{j=0}^m \binom{m}{j} p^j q^{m-j} \quad \begin{array}{l} j := k - 1 \\ m := n - 1 \end{array} \\ &= np \end{aligned}$$



Expectation

- Example
 - **Expectation of a Poisson random variable**
 - Consider random arrivals/hits occurring at a constant average rate $\lambda > 0$, i.e., λ arrivals/hits (typically) per unit time

$$\begin{aligned} E(X) &= \sum_{k \geq 0} k \frac{1}{k!} \lambda^k e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{k \geq 1} \frac{1}{(k-1)!} \lambda^{k-1} \\ &= \lambda e^{-\lambda} \sum_{j \geq 0} \frac{\lambda^j}{j!} \\ &= \lambda e^{-\lambda} e^\lambda \\ &= \lambda \end{aligned}$$



- This gives meaning to parameter λ as average number of arrivals in unit time

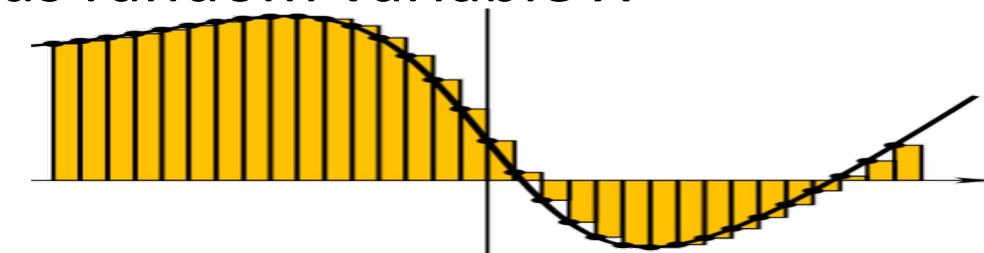
Expectation

- Definition:

Expectation of a Continuous Random variable: $E[X] := \int_{-\infty}^{\infty} x P(x) dx$

- Frequentist interpretation of probabilities and expectation

- If a random experiment underlying a continuous random variable X is repeated $N \rightarrow \infty$ times,
then,
for a tiny interval $[x, x + \Delta x]$,
the proportion of time X takes values within interval is approximately $P(x)\Delta x$
- So, in $N \rightarrow \infty$ experiments,
number of times we will get X within $[x_i, x_i + \Delta x]$ is approximately $N.P(x_i)\Delta x$
- So, across all $N \rightarrow \infty$ experiments,
arithmetic average of all observed values is
approximately $(1/N) \sum_i (x_i) (N.P(x_i)\Delta x)$
- In the limit that $\Delta x \rightarrow 0$, this average $\rightarrow E[X]$

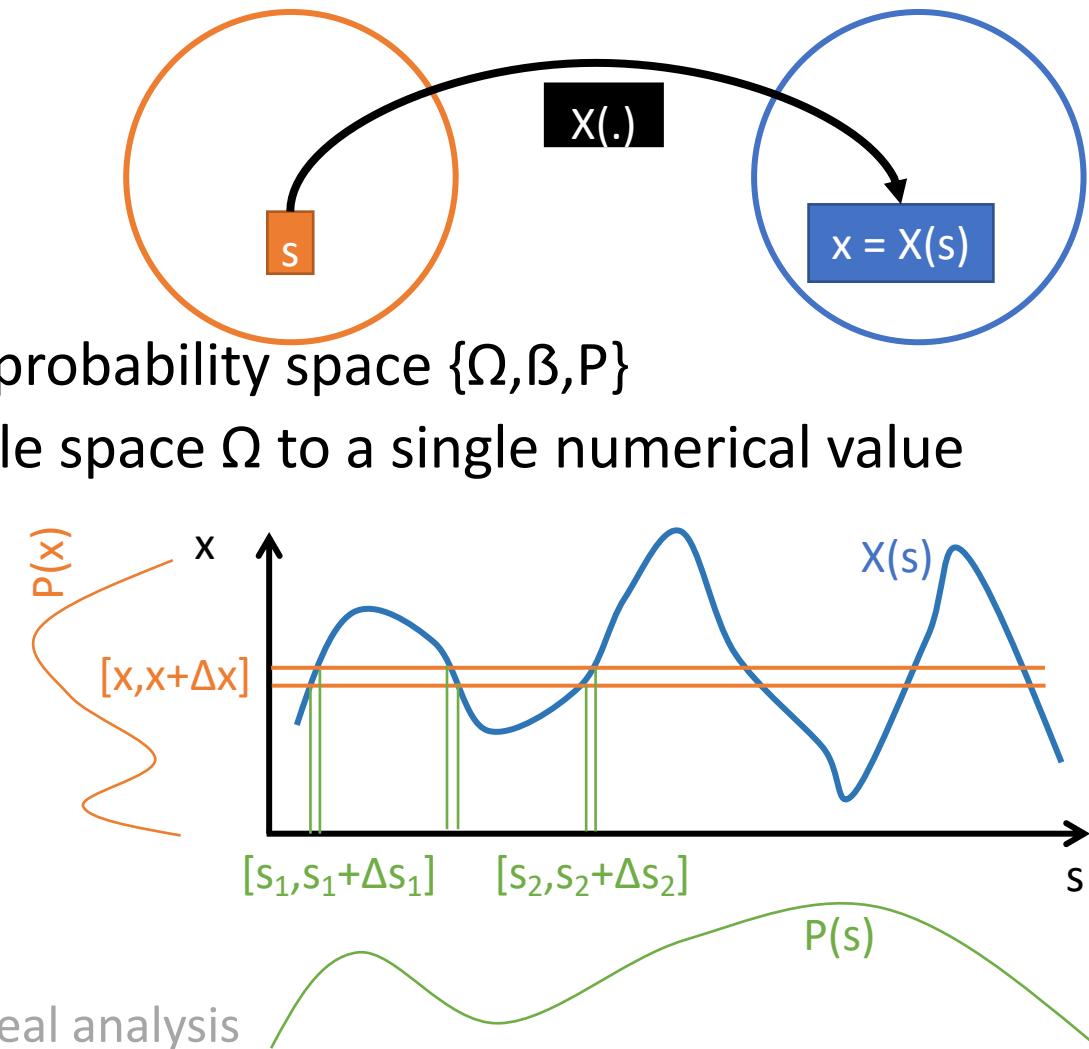


Expectation

- Another Formulation of Expectation

- Recall:

- Random variable X is a function defined on a probability space $\{\Omega, \mathcal{B}, P\}$
 - Function $X: \Omega \rightarrow \mathbb{R}$, maps each element in sample space Ω to a single numerical value belonging to the set of real numbers
 - $E[X] := \int_{-\infty}^{\infty} xP(x)dx = \int_{-\infty}^{\infty} X(s)P(s)ds$
 - Intuition remains the same as in the discrete case
 - Using probability-mass conservation:
 $P(x)\Delta x$ is approximated by $P(s_1)\Delta s_1 + P(s_2)\Delta s_2 + \dots$
 - Thus, $x.P(x)\Delta x$ is approximated by
 $X(s_1).P(s_1)\Delta s_1 + X(s_2).P(s_2)\Delta s_2 + \dots$
 - A more rigorous proof needs advanced results in real analysis



$$E[X] := \sum_i x_i p(X = x_i) = \sum_i x_i \left(\sum_{s \in \Omega: X(s)=x_i} P(s) \right) = \sum_{s \in \Omega} X(s)P(s)$$

Expectation

- Mean as the center of mass

- By definition,

$$\text{mean } m := E[X] := \int_x xP(x)dx$$

- Thus, $\int_x (x - m)P(x)dx = 0$

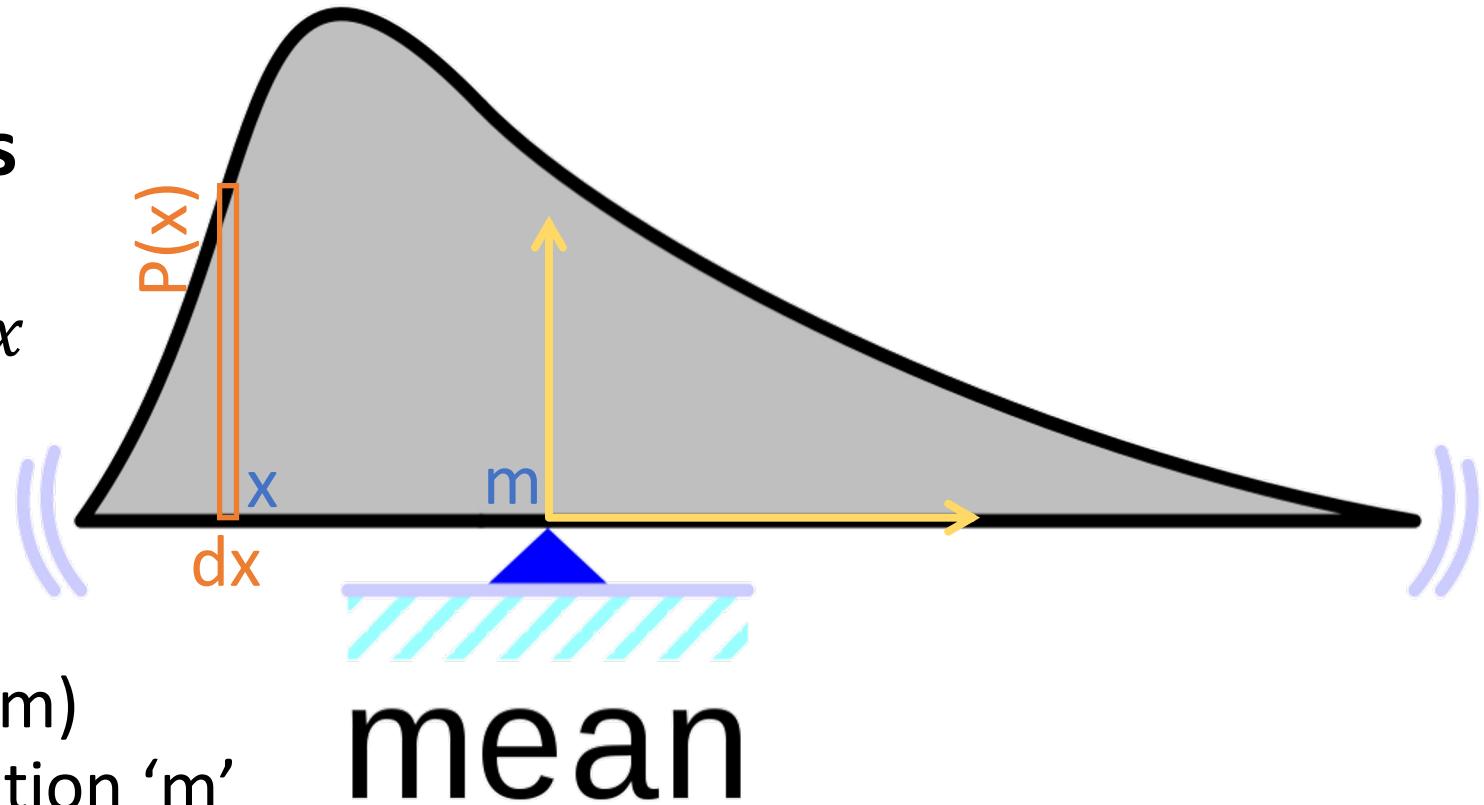
- Mass $P(x)dx$

placed around location 'x'

applies a torque $\propto P(x)dx.(x-m)$

at the fulcrum placed at location 'm'

- Because the integral $\int_x (x - m)P(x)dx$ is zero,
the net torque around the fulcrum 'm' is zero
- Hence, 'm' is the center of mass

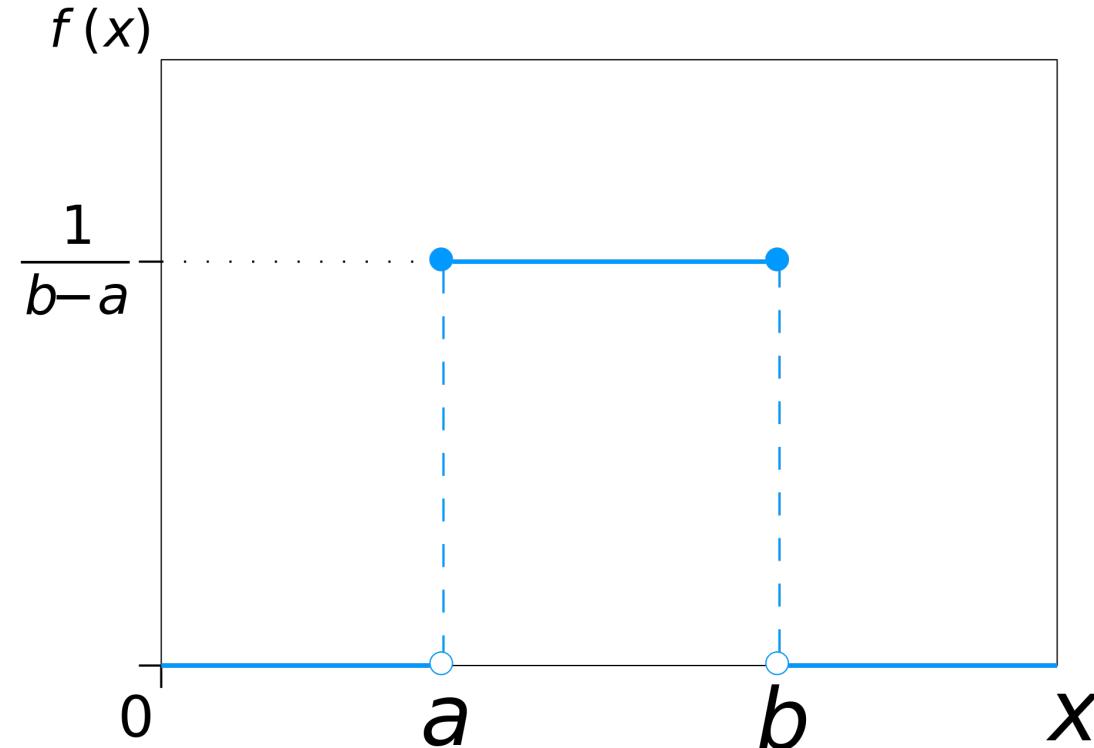


Expectation

- Example

- **Expectation of a uniform random variable (continuous case)**

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x) dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{2(b-a)} \left[x^2 \right]_a^b \\ &= \frac{b^2 - a^2}{2(b-a)} \\ &= \frac{b+a}{2} \end{aligned}$$



Expectation

- Example

- Expectation of an exponential random variable

- Consider random arrivals/hits occurring at a constant average rate $\lambda > 0$

- Define $\beta := 1/\lambda$

$$E(X) = \int_0^\infty x \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right) dx$$

$$= \beta \int_0^\infty u \exp(-u) du$$

$$u = \frac{x}{\beta}$$

$$= \left[-\beta(u + 1) \exp(-u) \right]_0^\infty$$

$$= \beta$$

- This gives meaning to parameter β as average inter-arrival time
 - Larger arrival/hit rate leads to lesser inter-arrival time

PDF

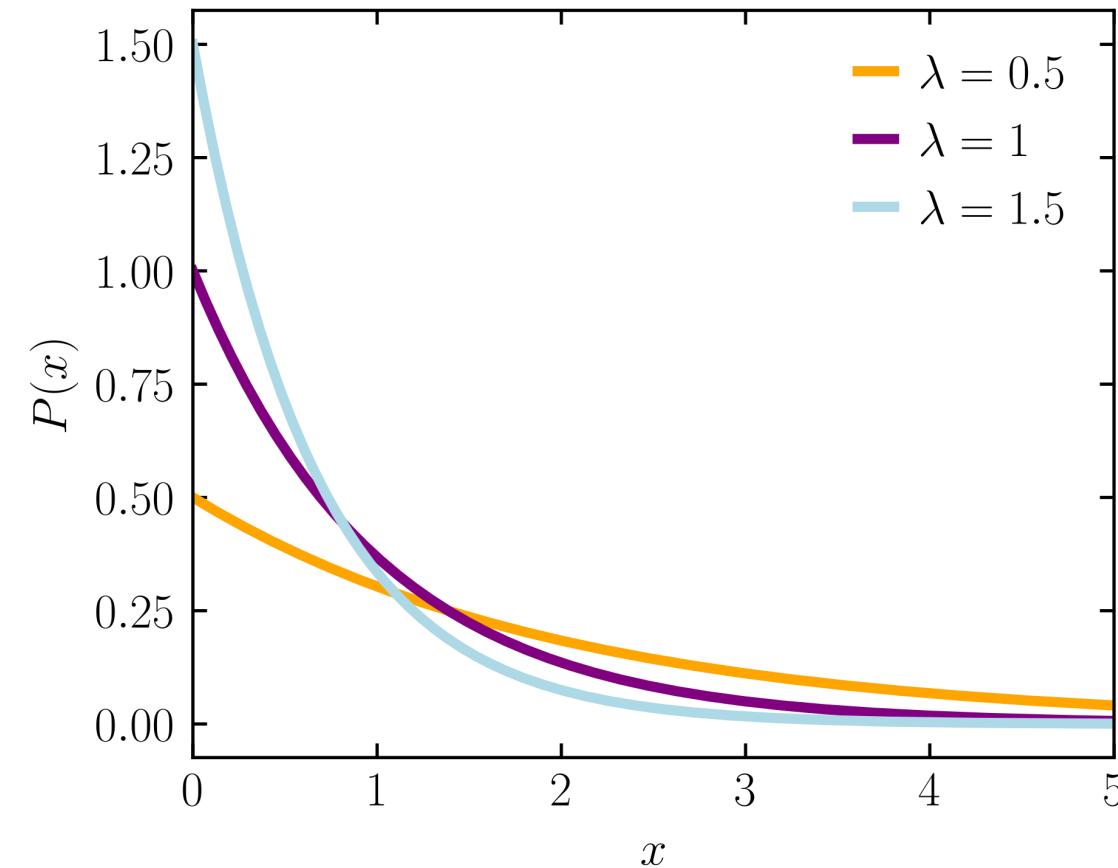
$$P(x) = 0, \text{ for all } x < 0$$

$$P(x) = \lambda \exp(-\lambda x), \forall x \geq 0$$

CDF

$$f(x) = 0, \text{ for all } x < 0$$

$$f(x) = 1 - \exp(-\lambda x), \forall x \geq 0$$



Expectation

- Example

- Expectation of a Gaussian random variable

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

$$= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu) \exp(-t^2) dt$$

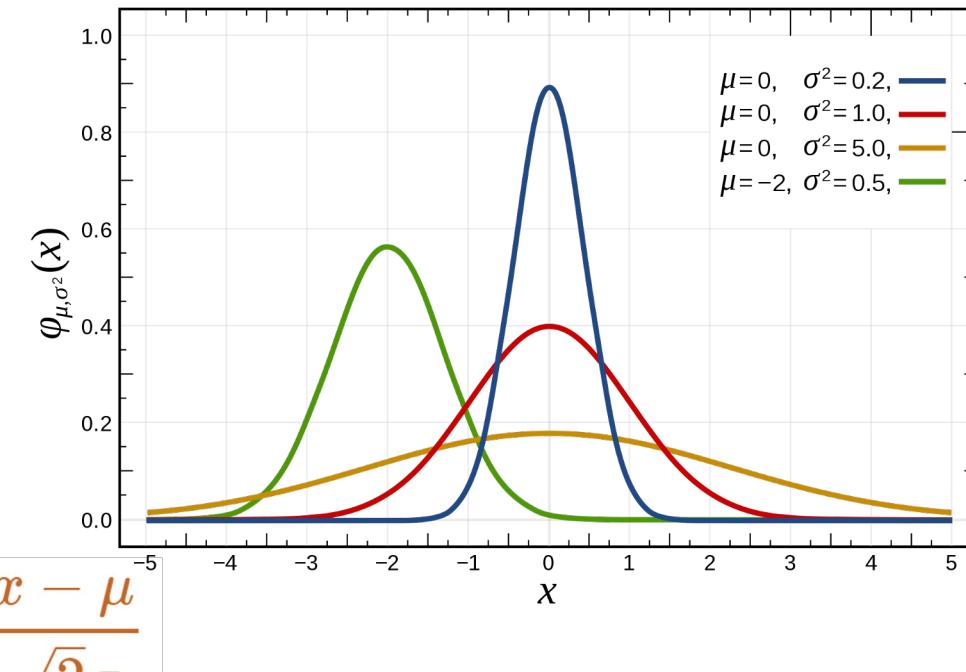
$$t = \frac{x - \mu}{\sqrt{2}\sigma}$$

$$= \frac{1}{\sqrt{\pi}} \left(\sqrt{2}\sigma \int_{-\infty}^{\infty} t \exp(-t^2) dt + \mu \int_{-\infty}^{\infty} \exp(-t^2) dt \right)$$

$$= \frac{1}{\sqrt{\pi}} \left(\sqrt{2}\sigma \left[-\frac{1}{2} \exp(-t^2) \right]_{-\infty}^{\infty} + \mu \sqrt{\pi} \right)$$

$$= \frac{\mu\sqrt{\pi}}{\sqrt{\pi}}$$

$$= \mu$$



Expectation

- Example

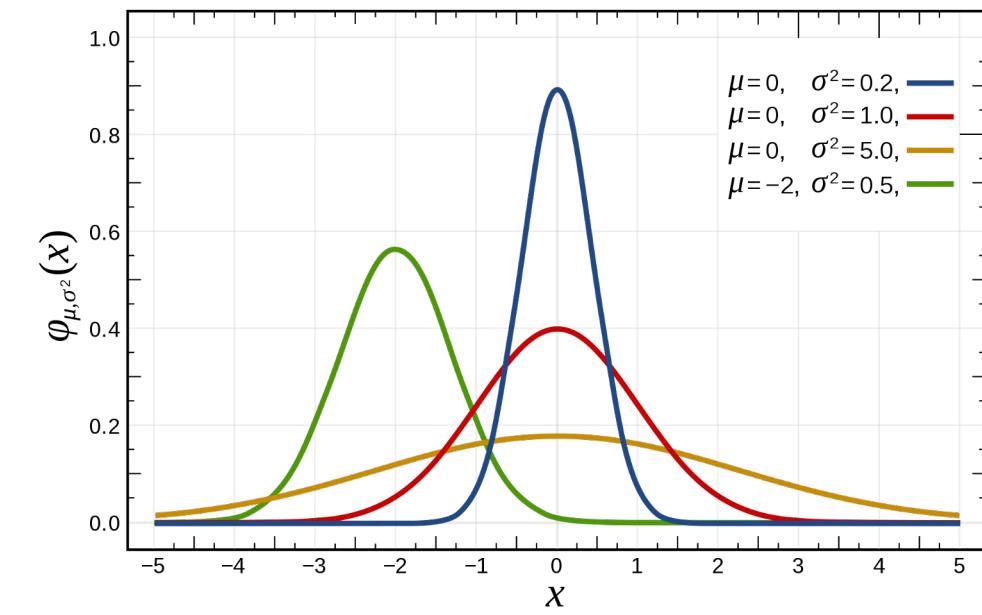
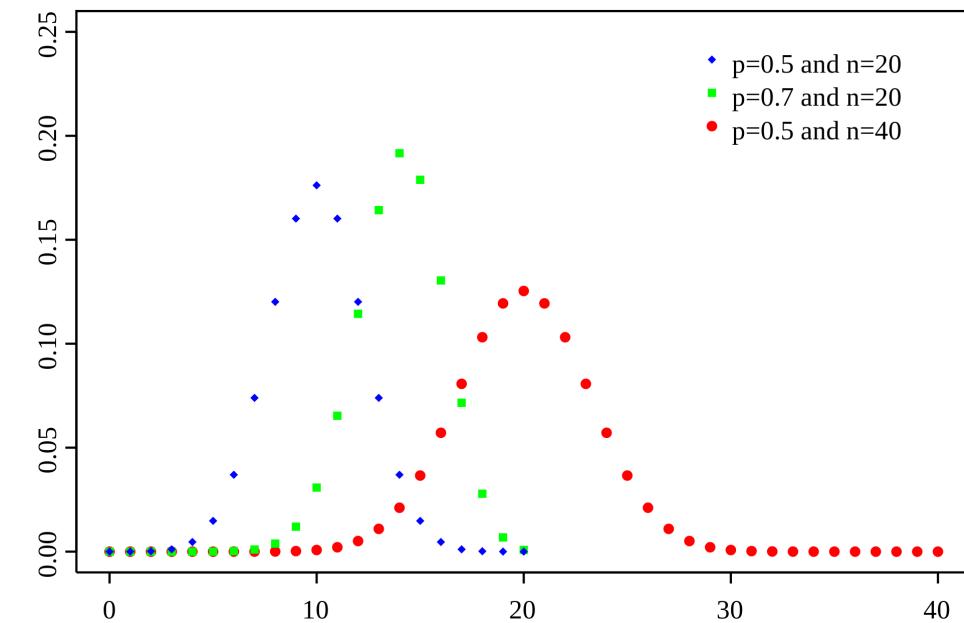
- Expectation of a limiting case of binomial
- As n tends to infinity,

binomial
$$f(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

tends to a

“Gaussian” form
$$\frac{1}{\sqrt{2\pi npq}} e^{-(x-np)^2/2npq}$$

- Gaussian expectation $\mu (=np \text{ here})$ is consistent with binomial expectation np



Expectation

- **Linearity of Expectation**

- For both discrete and continuous random variables
- For random variables X and Y having a joint probability space (Ω, \mathcal{B}, P) , the following rules hold:

- $E[X + Y] = E[X] + E[Y]$

- Either $E[X + Y] = \sum_{s \in \Omega} (X(s) + Y(s))P(s) = E[X] + E[Y]$

- Or LHS = $\int_x \int_y (x + y)P(x, y)dx dy = \int_x x \int_y P(x, y)dy dx + \int_y y \int_x P(x, y)dx dy = \text{RHS}$

- $E[X + c] = E[X] + c$, where 'c' is a constant

$$E[X + c] = \sum_{s \in \Omega} (X(s) + c)P(s) = \sum_{s \in \Omega} X(s)P(s) + c \sum_{s \in \Omega} P(s) = E[X] + c$$

- $E[aX] = aE[X]$, where 'a' is a scalar constant

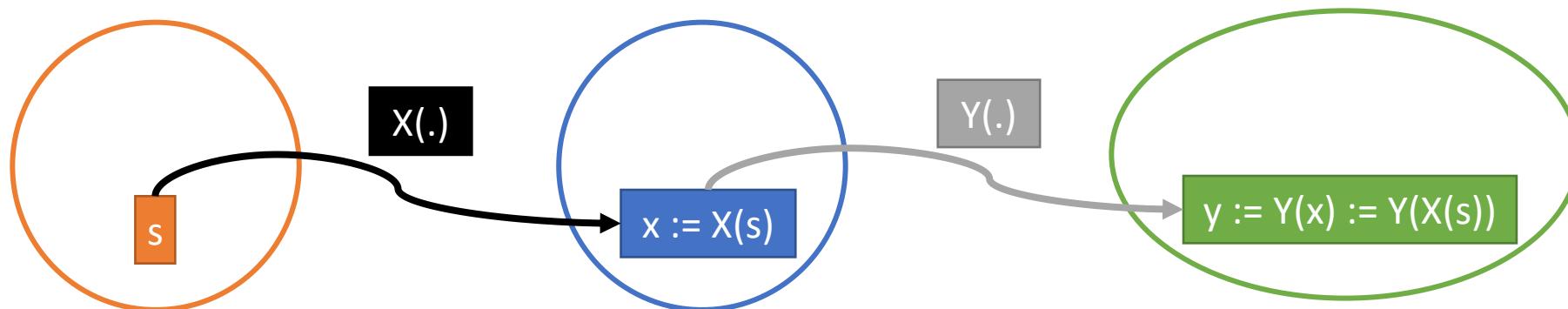
$$E[aX] = \sum_{s \in \Omega} aX(s)P(s) = aE[X]$$

- This generalizes to:

$$E[\sum_i a_i X_i] = \sum_i a_i E[X_i]$$

Expectation

- **Expectation of a “function of a random variable”**
 - Let us define values $y := Y(x)$, or “ $Y(\cdot)$ is a function of the random variable X ”



- **Discrete random variable:** $E[Y(X)] := E_{P(X)}[Y(X)] := \sum_{x_i} Y(x_i)P(x_i)$
- **Continuous random variable:** $E[Y(X)] := E_{P(X)}[Y(X)] := \int_x Y(x)P(x)dx$
- **Property:**
 - Just as $E_{P(S)}[X(S)] = E_{P(X)}[X]$, ...
 - ... we get $E_{P(X)}[Y(X)] = E_{P(Y)}[Y]$

$$E[X] := \sum_i x_i p(X = x_i) = \sum_i x_i \left(\sum_{s \in \Omega: X(s) = x_i} P(s) \right) = \sum_{s \in \Omega} X(s)P(s)$$

Expectation

- Expectation of a function of multiple random variables

- **Definition:** When we have multiple random variables X_1, \dots, X_n with a joint PMF/PDF $P(X_1, \dots, X_n)$ and a function of the multiple random variables $g(X_1, \dots, X_n)$, then we define the expectation of $g(X_1, \dots, X_n)$ as:

$$E[g(X_1, \dots, X_n)] := \sum_{x_1, \dots, x_n} g(x_1, \dots, x_n) P(X_1 = x_1, \dots, X_n = x_n)$$

or

$$E[g(X_1, \dots, X_n)] := \int_{x_1, \dots, x_n} g(x_1, \dots, x_n) P(x_1, \dots, x_n) dx_1 \dots dx_n$$

- If X and Y are independent, then $E[XY] = E[X] E[Y]$

- Proof:

$$\sum_{x,y} xy P(X = x, Y = y) = \sum_{x,y} xy P(X = x)P(Y = y) = \sum_x x P(X = x) \sum_y y P(Y = y)$$

Expectation

- **Tail-sum formula**

- Let X be a **discrete** random variable taking **values** in set of **natural numbers**

- Then, $E(X) = \sum_{k=1}^{\infty} \Pr(X \geq k)$

P(x=1)

P(x=2) P(x=2)

P(x=3) P(x=3) P(x=3)

P(x=4) P(x=4) P(x=4) P(x=4)

...

- Proof: $E(X) = \sum_{x=1}^{\infty} x \Pr(X = x)$

Sum over rows (row number = x)

$$= \sum_{x=1}^{\infty} \sum_{k=1}^x \Pr(X = x)$$

$$= \sum_{k=1}^{\infty} \sum_{x=k}^{\infty} \Pr(X = x)$$

Sum over columns (column number = k)

$$= \sum_{k=1}^{\infty} \Pr(X \geq k)$$

Expectation

- Tail-sum formula

- Let X be a **continuous** random variable taking **non-negative** values

- Notation: For random variable X , PDF is $f_X(\cdot)$ and CDF is $F_X(\cdot)$

- Then,

$$E(X) = \int_0^\infty (1 - F_X(x)) dx$$

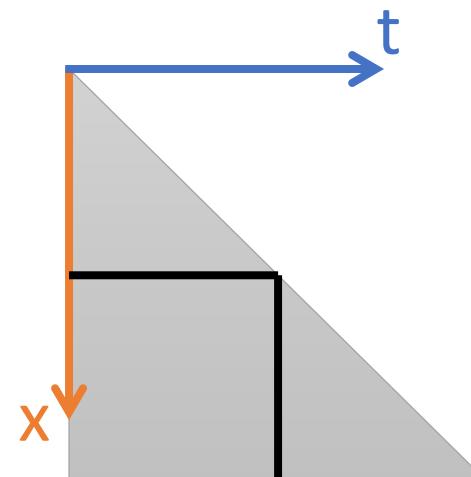
- Proof: $E(X) = \int_0^\infty x f_X(x) dx$

$$= \int_0^\infty \int_0^x f_X(x) dt dx$$

$$= \int_0^\infty \int_t^\infty f_X(x) dx dt$$

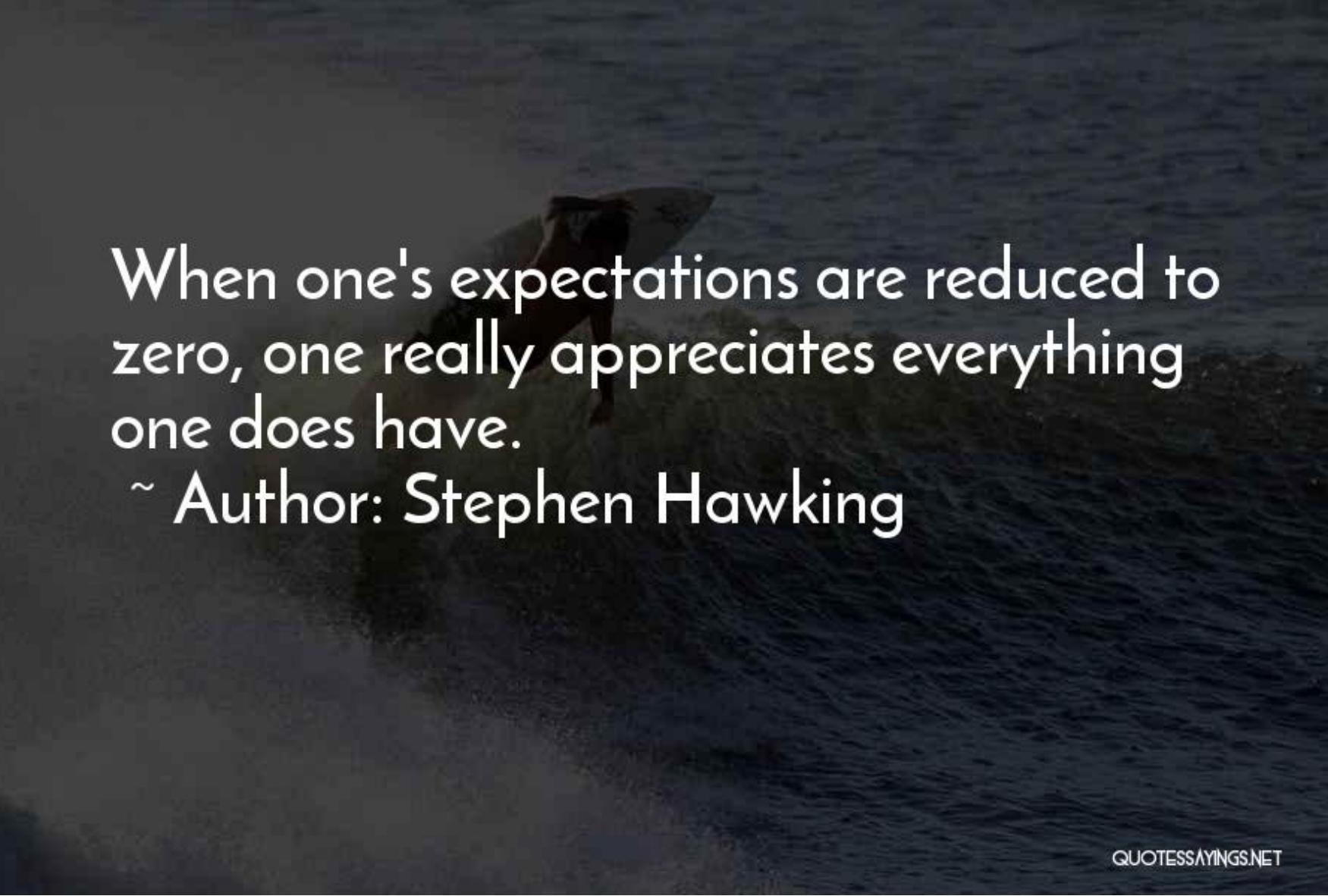
$$= \int_0^\infty \Pr(X > t) dt$$

$$= \int_0^\infty (1 - F_X(t)) dt$$



Expectation in Life

- Action without expectation → Happiness [Indian Philosophy]

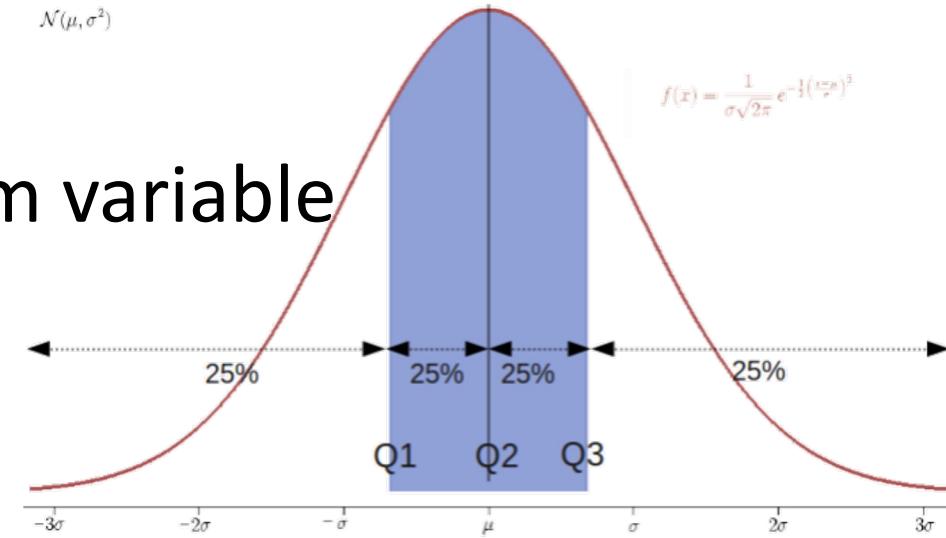


When one's expectations are reduced to zero, one really appreciates everything one does have.

~ Author: Stephen Hawking

Quantile, Quartile

- Definition: For a **discrete/continuous** random variable with a PMF/PDF $P(\cdot)$, the q-th **quantile** (where $0 < q < 1$) is any real number ' x_q ' such that $P(X \leq x_q) \geq q$ and $P(X \geq x_q) \geq 1-q$



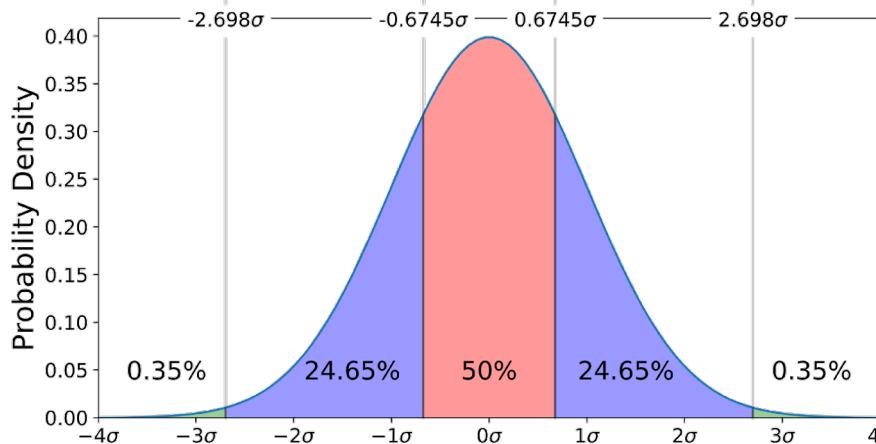
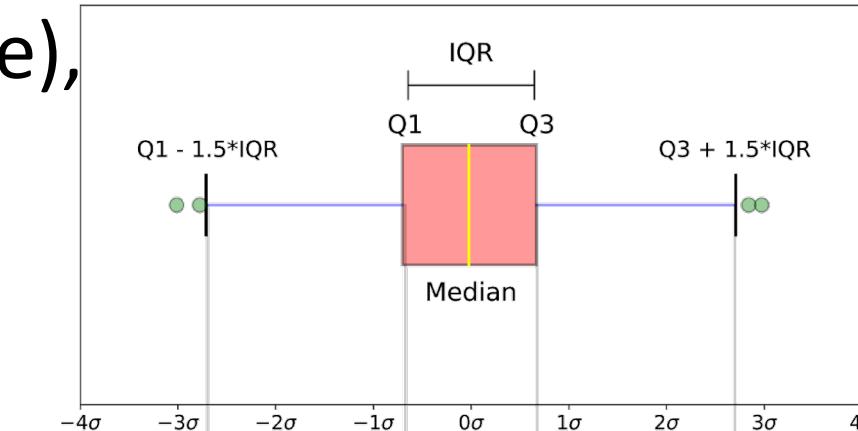
- Quartiles: $q = 0.25$ (1st quartile), $q = 0.5$ (2nd), $q = 0.75$ (3rd)

- Percentiles

- $q=0.25 \rightarrow 25^{\text{th}}$ percentile

- Box plot,
box-and-whisker plot

- Inter-Quartile Range (IQR)



Quantile, Median

- Definition:

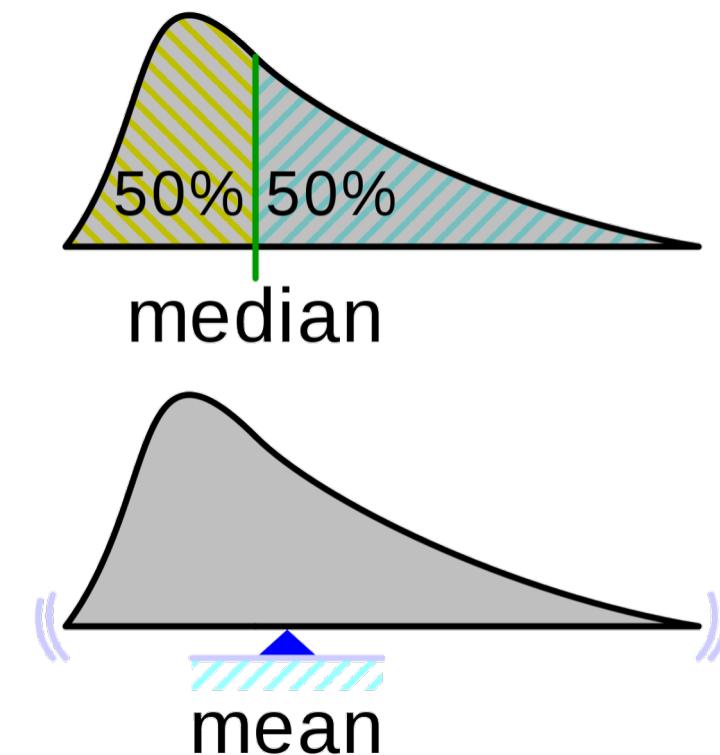
For a **discrete/continuous** random variable with a PMF/PDF $P(\cdot)$, the **median** is any real number ‘ m ’ such that $P(X \leq m) \geq 0.5$ and $P(X \geq m) \geq 0.5$

- Median = second quartile

- Definition:

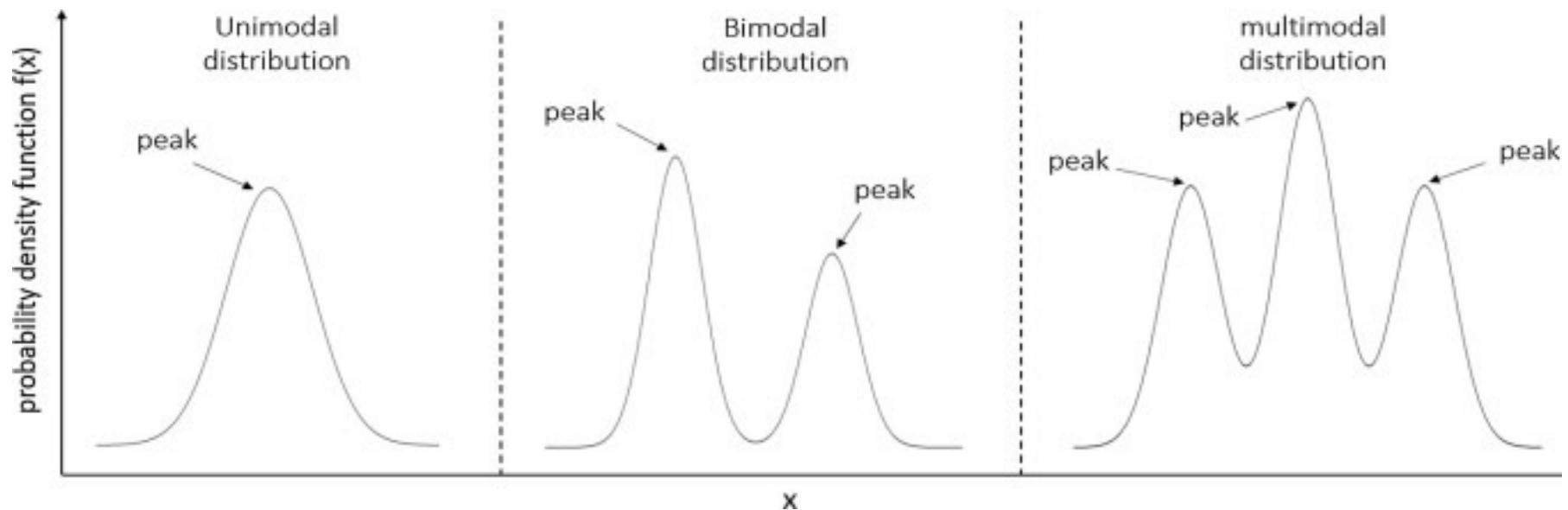
For a **continuous** random variable with a PDF $P(\cdot)$, the median is any real number ‘ m ’ such that $P(X \leq m) = P(X > m)$

- CDF $f_X(m) = 0.5$
- A PDF can be associated with multiple medians



Mode

- For **discrete X**
 - Mode m is a value for which the PMF value $P(X=m)$ is maximum
 - A PMF can have multiple modes
- For **continuous X**
 - Mode ‘ m ’ is any **local** maximum of the PDF $P(\cdot)$
 - A PDF can have multiple modes
 - **Unimodal PDF** = A PDF having only 1 local maximum
 - **Bimodal PDF**:
2 local maxima
 - **Multimodal PDF**:
2 or more
local maxima



Mean, Median, Mode

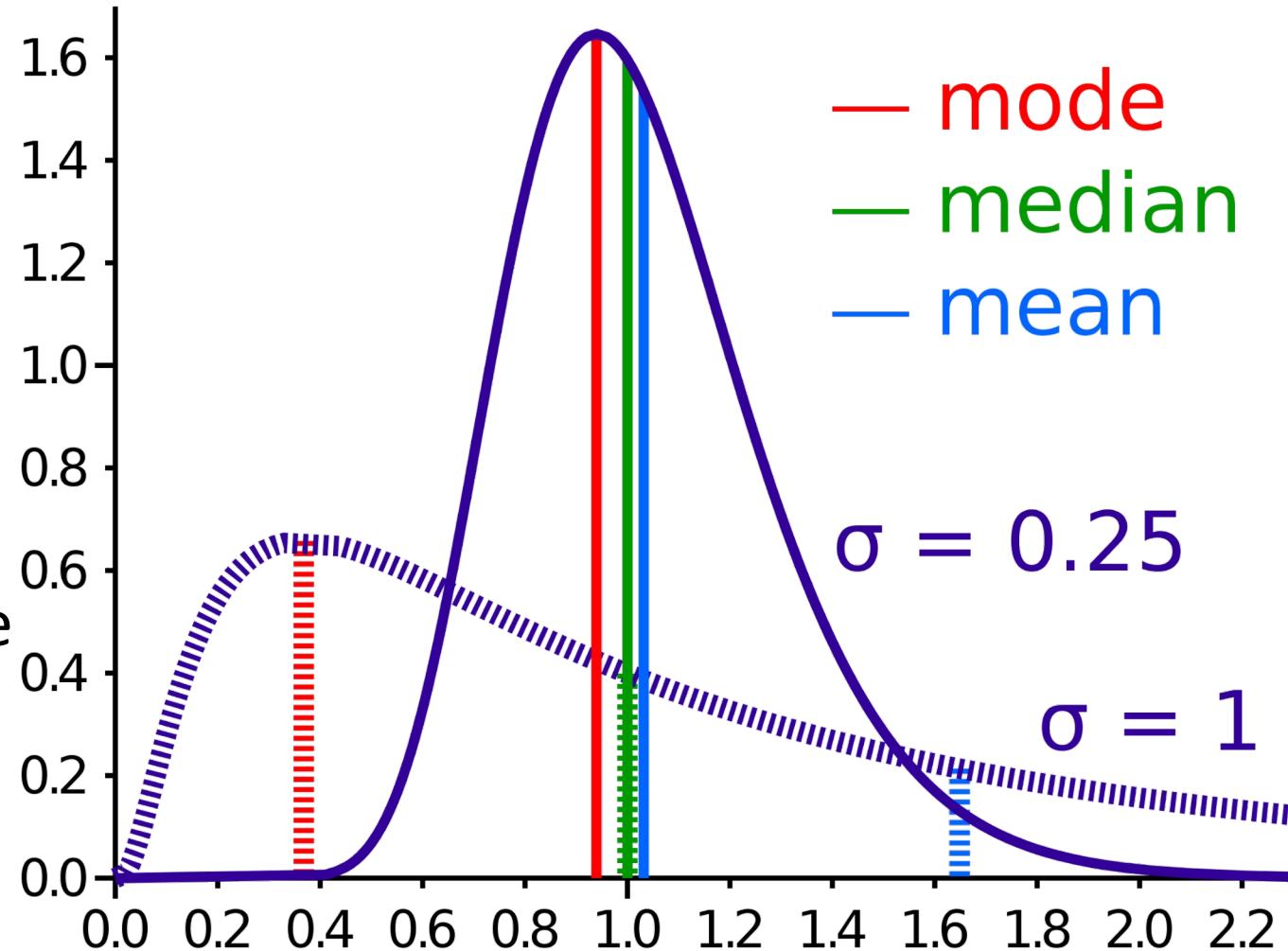
- For **continuous X**, for unimodal and symmetric distributions,
mode = mean = median

- Assuming symmetry around mode,
mass on left of mode = mass on right of mode

- So, mode = median

- Assuming symmetry around mode,
every $P(x)dx$ mass on left of mode
is matched by
 $P(x)dx$ mass on right of mode

- So, mode = mean



Variance

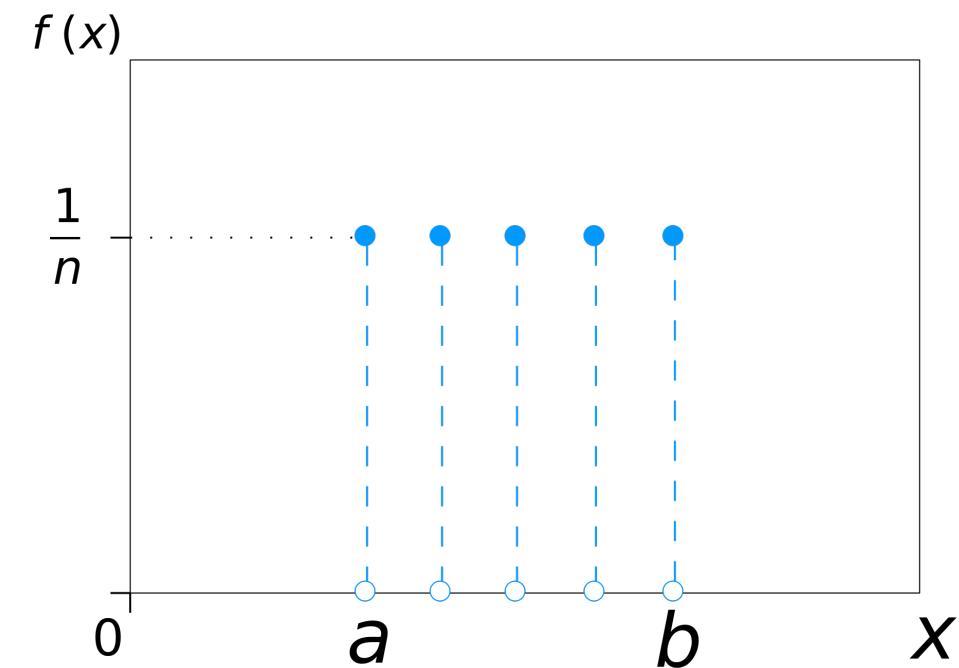
- **Definition:** $\text{Var}(X) := E[(X-E[X])^2]$
- A measure of the spread of the mass (in PMF or PDF) around the mean
- Property: Variance is always non-negative
- Property: $\text{Var}(X) = E[X^2] - (E[X])^2$
 - Proof: LHS =
 $E[(X-E[X])^2]$
= $E[X^2 + (E[X])^2 - 2.X.E[X]]$
= $E[X^2] + (E[X])^2 - 2(E[X])^2$
= $E[X^2] - (E[X])^2 = \text{RHS}$
- **Definition: Standard deviation** is the square root of the variance
- Units of variance = square of units of values taken by random variable
- Units of standard deviation = units of values taken by random variable

Variance

- **Variance of a Uniform Random Variable**

- Discrete case

- X has uniform distribution over n integers $\{a, a+1, \dots, b\}$
- Here, $n = b-a+1$
- Variance = $(n^2 - 1) / 12$

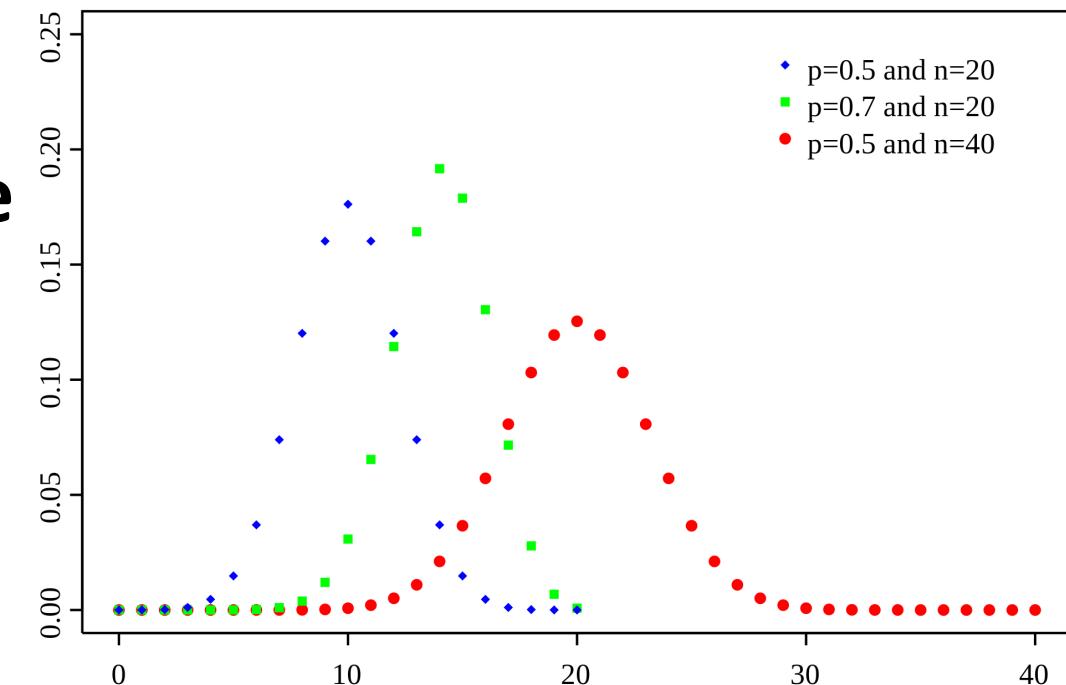


Variance

- **Variance of a Binomial Random Variable**

- $\text{Var}(X) = E[X^2] - (E[X])^2$, where $E[X] = np$

$$\begin{aligned} E(X^2) &= \sum_{k=0}^n k^2 \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=0}^n kn \binom{n-1}{k-1} p^k q^{n-k} \\ &= np \sum_{k=1}^n k \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} \\ &= np \sum_{j=0}^m (j+1) \binom{m}{j} p^j q^{m-j} \\ &= np \left(\sum_{j=0}^m j \binom{m}{j} p^j q^{m-j} + \sum_{j=0}^m \binom{m}{j} p^j q^{m-j} \right) \end{aligned}$$



$$j = k - 1, m = n - 1$$

Variance

- Variance of a Binomial Random Variable

- $\text{Var}(X) = E[X^2] - (E[X])^2$, where $E[X] = np$

- So, $E[X^2]$

$$= np(mp + 1)$$

$$= np((n-1)p + 1)$$

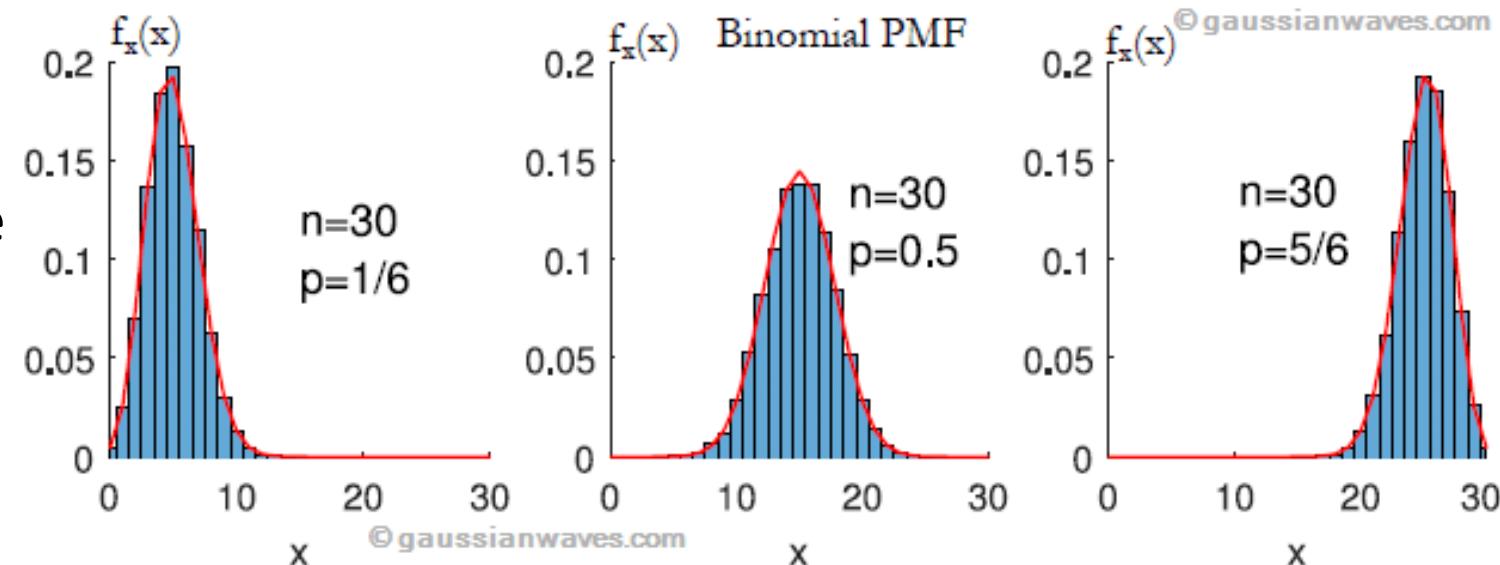
$$= (np)^2 + np(1-p)$$

- Thus, $\text{Var}(X) = np(1-p) = npq$

- Interpretation

- When $p=0$ or $p=1$,
then $\text{Var}(X) = 0$,
which is the minimum possible
- When $p=q=0.5$,
then $\text{Var}(X)$ is maximized

$$= np \left(\sum_{j=0}^m j \binom{m}{j} p^j q^{m-j} + \sum_{j=0}^m \binom{m}{j} p^j q^{m-j} \right)$$

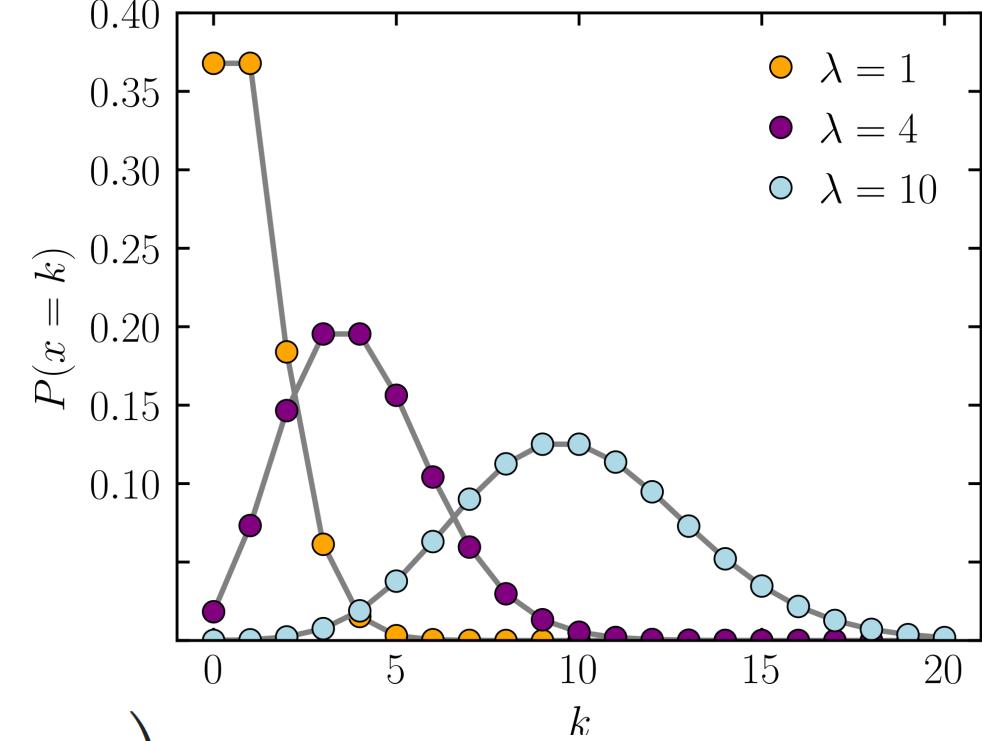


Variance

- **Variance of a Poisson Random Variable**

- $\text{Var}(X) = E[X^2] - (E[X])^2$, where $E[X] = \lambda$

$$\begin{aligned} E(X^2) &= \sum_{k \geq 0} k^2 \frac{1}{k!} \lambda^k e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{k \geq 1} k \frac{1}{(k-1)!} \lambda^{k-1} \\ &= \lambda e^{-\lambda} \left(\sum_{k \geq 1} (k-1) \frac{1}{(k-1)!} \lambda^{k-1} + \sum_{k \geq 1} \frac{1}{(k-1)!} \lambda^{k-1} \right) \\ &= \lambda e^{-\lambda} \left(\lambda \sum_{k \geq 2} \frac{1}{(k-2)!} \lambda^{k-2} + \sum_{k \geq 1} \frac{1}{(k-1)!} \lambda^{k-1} \right) \\ &= \lambda e^{-\lambda} \left(\lambda \sum_{i \geq 0} \frac{1}{i!} \lambda^i + \sum_{j \geq 0} \frac{1}{j!} \lambda^j \right) \end{aligned}$$



$$i = k-2, j = k-1$$

Variance

- Variance of a Poisson Random Variable

- $\text{Var}(X) = E[X^2] - (E[X])^2$, where $E[X] = \lambda$

- So, $E[X^2]$

$$= \lambda (\lambda \cdot 1 + 1)$$

$$= \lambda^2 + \lambda$$

- Thus, $\text{Var}(X) = \lambda$

- Interpretation

- Mean of Poisson random variable was also λ

- Standard deviation of Poisson random variable is $\lambda^{0.5}$

- As mean increases, so does variance (and standard deviation)

- When mean increase by factor of N (i.e., N time larger signal = number of arrivals/hits), then the standard deviation (spread) increases only by a factor of $N^{0.5}$

- As N increases,
then variability in number of arrivals/hits, relative to average arrival/hit rate, decreases

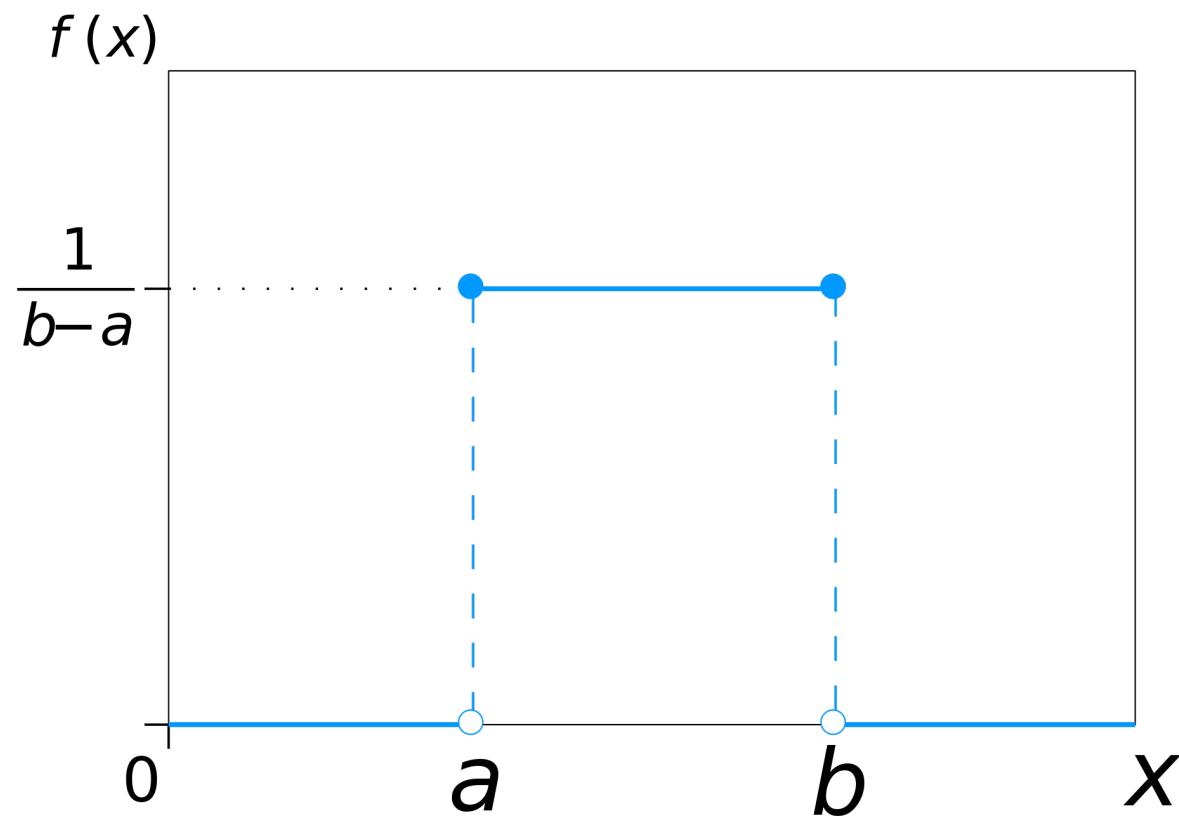
$$= \lambda e^{-\lambda} \left(\lambda \sum_{i \geq 0} \frac{1}{i!} \lambda^i + \sum_{j \geq 0} \frac{1}{j!} \lambda^j \right)$$

Variance

- **Variance of a Uniform Random Variable**

- Continuous case

- X has uniform distribution over $[a,b]$
- Variance = $(b - a)^2 / 12$



Variance

• Variance of a Exponential Random Variable

- $\text{Var}(X) = E[X^2] - (E[X])^2$, where $E[X] = \beta := 1/\lambda$

$$E(X^2) = \int_{x \in \Omega_X} x^2 f_X(x) dx$$

$$= \int_0^\infty x^2 \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right) dx$$

$$= \left[-x^2 \exp\left(-\frac{x}{\beta}\right) \right]_0^\infty + \int_0^\infty 2x \exp\left(-\frac{x}{\beta}\right) dx$$

$$= 0 + 2\beta \int_0^\infty x \frac{1}{\beta} \exp\left(-\frac{x}{\beta}\right) dx$$

$$= 2\beta E(X)$$

$$= 2\beta^2$$

- So, $\text{Var}(X) = \beta^2$. So, $\beta = E[X] = SD(X)$; unlike Poisson.

PDF

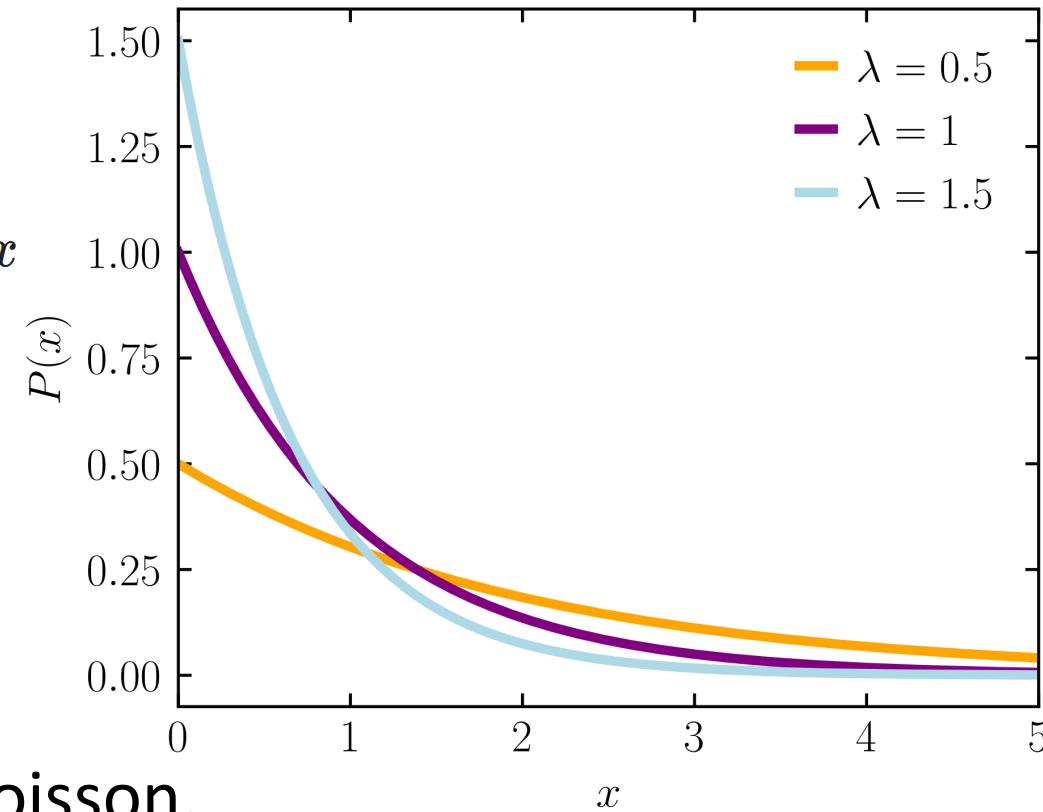
$$P(x) = 0, \text{ for all } x < 0$$

$$P(x) = \lambda \exp(-\lambda x), \forall x \geq 0$$

CDF

$$f(x) = 0, \text{ for all } x < 0$$

$$f(x) = 1 - \exp(-\lambda x), \forall x \geq 0$$



Variance

- **Variance of a Gaussian Random Variable**

- $\text{Var}(X) = E[X^2] - (E[X])^2$, where $E[X] = \mu$

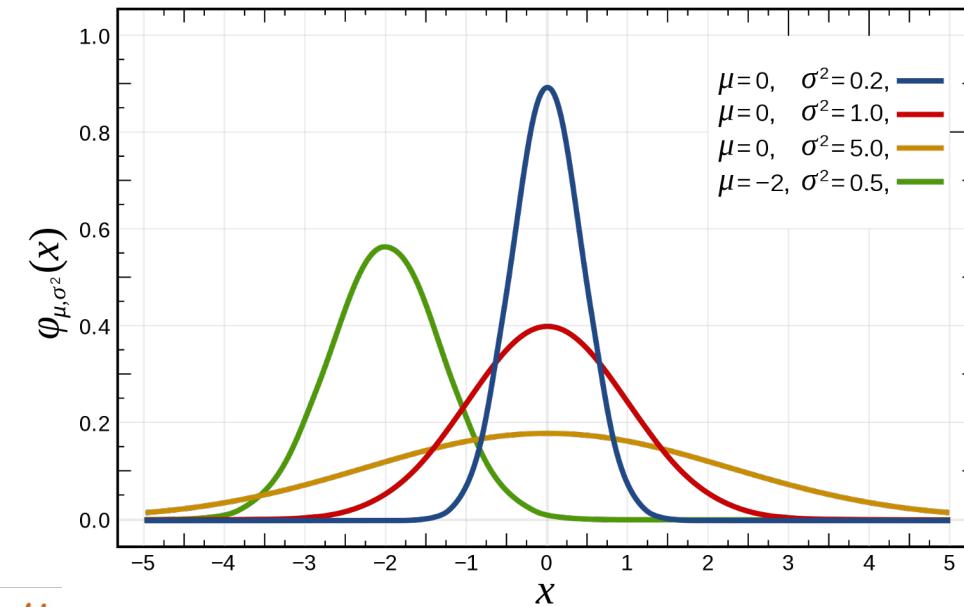
$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx - \mu^2$$

$$= \frac{\sqrt{2}\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sqrt{2}\sigma t + \mu)^2 \exp(-t^2) dt - \mu^2$$

$$= \frac{1}{\sqrt{\pi}} \left(2\sigma^2 \int_{-\infty}^{\infty} t^2 \exp(-t^2) dt + 2\sqrt{2}\sigma\mu \int_{-\infty}^{\infty} t \exp(-t^2) dt + \mu^2 \int_{-\infty}^{\infty} \exp(-t^2) dt \right) - \mu^2$$

$$= \frac{1}{\sqrt{\pi}} \left(2\sigma^2 \int_{-\infty}^{\infty} t^2 \exp(-t^2) dt + 2\sqrt{2}\sigma\mu \cdot 0 \right) + \mu^2 - \mu^2$$

$$t = \frac{x - \mu}{\sqrt{2}\sigma}$$

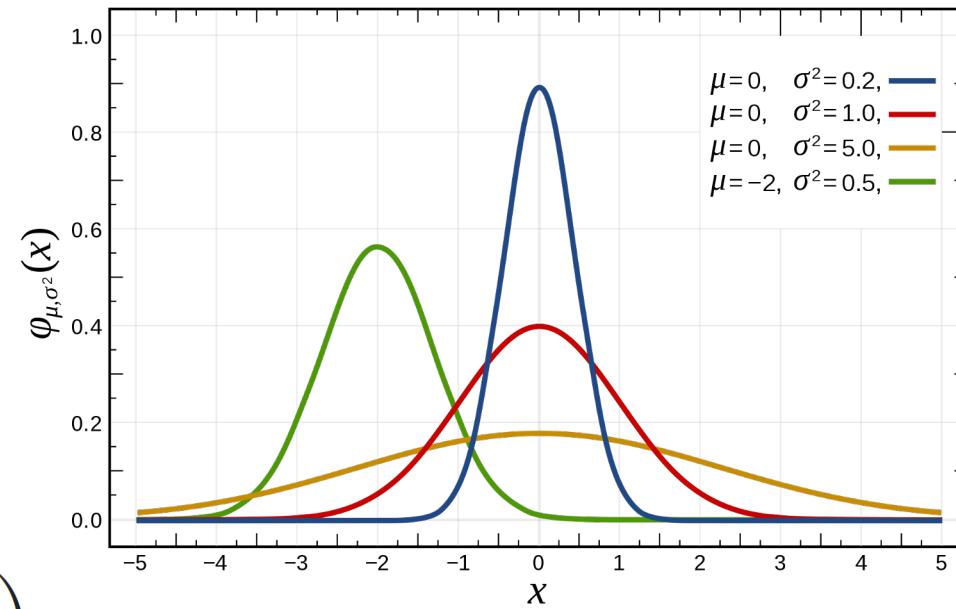


Variance

- Variance of a Gaussian Random Variable

- $\text{Var}(X) = E[X^2] - (E[X])^2$, where $E[X] = \mu$

$$\begin{aligned} &= \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 \exp(-t^2) dt \\ &\quad \boxed{t \cdot (t \exp(-t^2))} \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \left(\left[-\frac{t}{2} \exp(-t^2) \right]_{-\infty}^{\infty} + \frac{1}{2} \int_{-\infty}^{\infty} \exp(-t^2) dt \right) \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2} \int_{-\infty}^{\infty} \exp(-t^2) dt \\ &= \frac{2\sigma^2 \sqrt{\pi}}{2\sqrt{\pi}} \\ &= \sigma^2 \end{aligned}$$



Variance

- Example

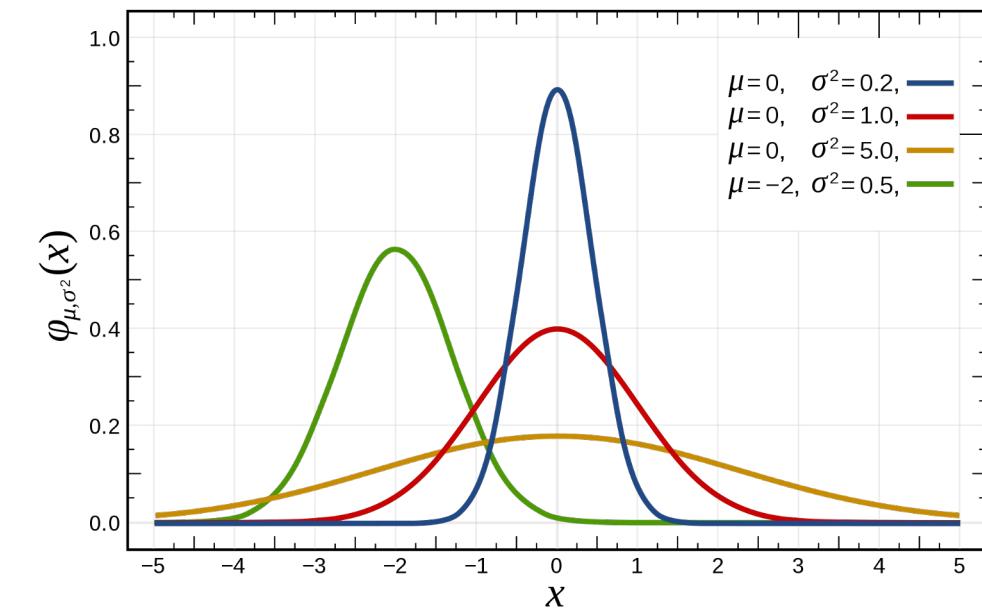
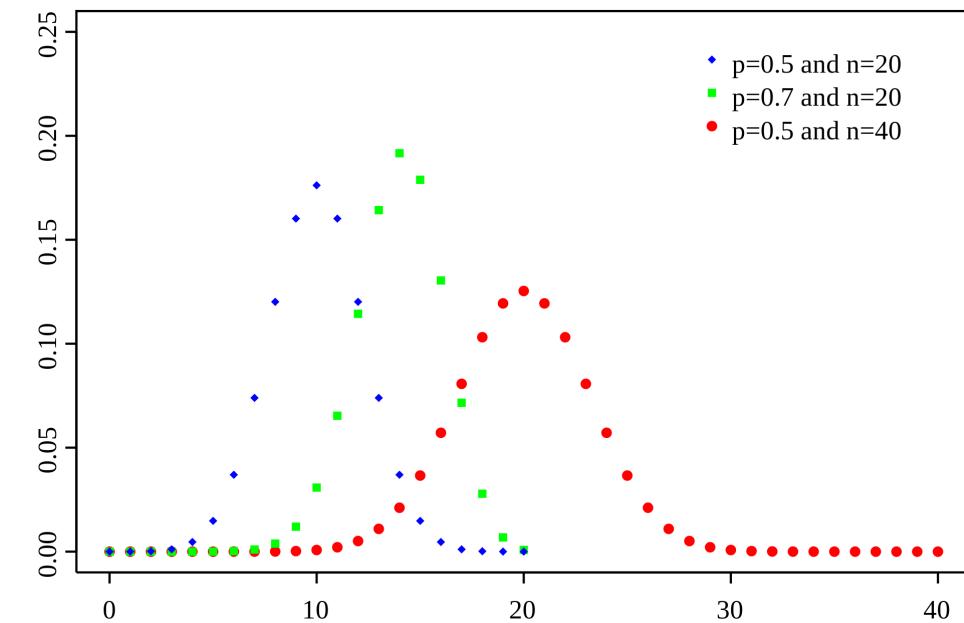
- Variance of a limiting case of binomial
- As n tends to infinity,

binomial
$$f(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

tends to

Gaussian
$$\frac{1}{\sqrt{2\pi npq}} e^{-(x-np)^2/2npq}$$

- Gaussian variance σ^2 ($= npq$ in this case) is consistent with binomial variance npq



Variance

- **Property: $\text{Var}(aX+c) = a^2\text{Var}(X)$**

- Adding a constant to a random variable doesn't change the variance (spread)
 - This only shifts the PDF/PMF
 - If $Y := X + c$, then $\text{Var}(Y) = \text{Var}(X)$
- If we scale a random variable by 'a', then the variance gets scaled by a^2
 - If $Y := aX$, then $\text{Var}(Y) = a^2\text{Var}(X)$

- Proof:

$$\begin{aligned}\text{Var}(aX + c) &= E((aX + c)^2) - E(aX + c)^2 \\&= E(a^2X^2 + 2acX + c^2) - (aE(X) + c)^2 \\&= a^2E(X^2) + 2acE(X) + c^2 - a^2E(X)^2 - 2acE(X) - c^2 \\&= a^2(E(X^2) - E(X)^2) \\&= a^2 \text{Var}(X)\end{aligned}$$

Variance

- **Property:** $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2(E[XY] - E[X]E[Y])$
- **Proof:**
$$\begin{aligned} \text{Var}(X + Y) &= E((X + Y)^2) - E(X + Y)^2 \\ &= E(X^2 + 2XY + Y^2) - (E(X) + E(Y))^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\ &= (E(X^2) - E(X)^2) + (E(Y^2) - E(Y)^2) + 2(E(XY) - E(X)E(Y)) \\ &= \text{Var}(X) + \text{Var}(Y) + 2(E(XY) - E(X)E(Y)) \end{aligned}$$
- If X and Y are independent,
then $E[XY] = E[X]E[Y]$, and so $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$
- If X,Y,Z are independent, then
 $\text{Var}(X+Y+Z) = \text{Var}(X+Y) + \text{Var}(Z) = \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z)$
- For independent random variables X_1, \dots, X_n ;
 $\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n)$

Markov's Inequality

- **Theorem:** Let X be a random variable with PDF $P(\cdot)$.

Let $u(\cdot)$ be an non-negative-valued function.

Let 'c' be a positive constant.

$$\text{Then, } P(u(X) \geq c) \leq E[u(X)] / c$$

- Proof:

- $E[u(X)] = \int_{x:u(x) \geq c} u(x) P(x) dx + \int_{x:u(x) < c} u(x) P(x) dx$

- Because $u(\cdot)$ takes non-negative values, each integral above is non-negative

- So, $E[u(X)] \geq \int_{x:u(x) \geq c} u(x) P(x) dx$

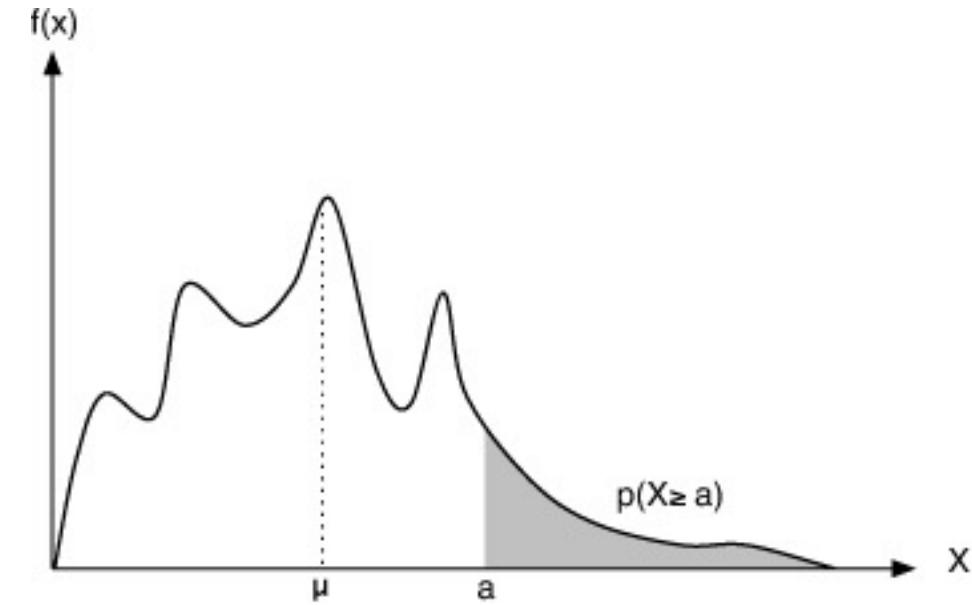
$$\geq c \int_{x:u(x) \geq c} P(x) dx$$

$$= c P(u(X) \geq c)$$

- Because $c > 0$, we get $E[u(X)]/c \geq P(u(X) \geq c)$

- Special case →

- X takes non-negative values & $u(x) := x$



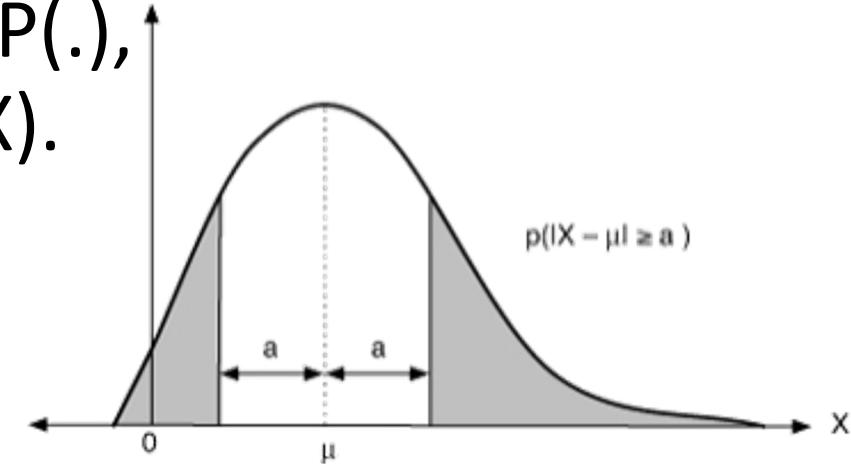
Chebyshev's Inequality

Markov's Inequality:
 $P(u(X) \geq c) \leq E[u(X)] / c$

- **Theorem:** Let X be a random variable with PDF $P(\cdot)$, finite expectation $E[X]$, and finite variance $\text{Var}(X)$. Then, $P(|X-E[X]| \geq a) \leq \text{Var}(X) / a^2$

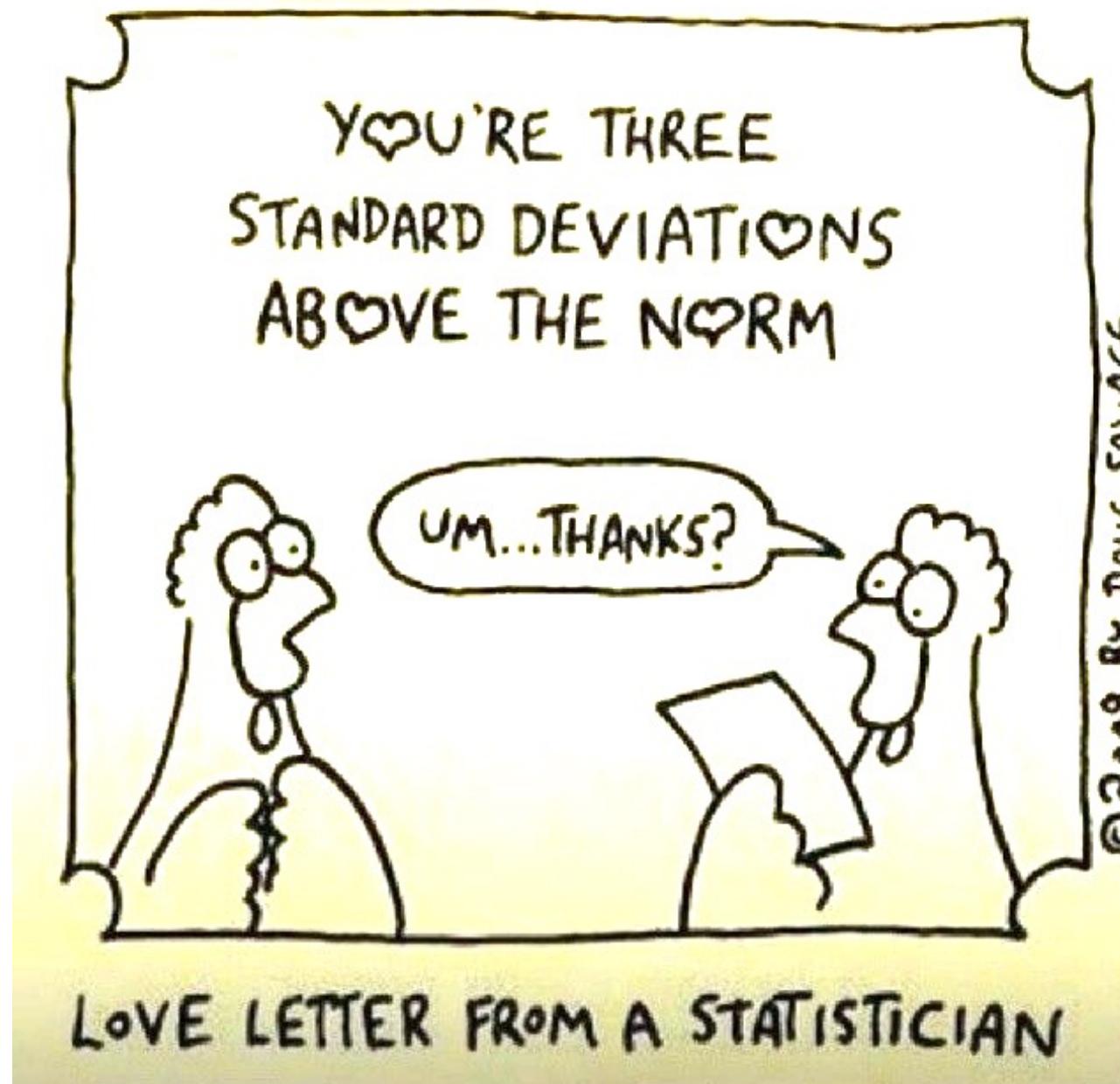
- **Proof:**

- Define random variable $u(X) := (X-E[X])^2$
- Then, by Markov's inequality, $P(u(X) \geq a^2) \leq E[u(X)] / a^2$
- LHS = $P(|X-E[X]| \geq a)$
- RHS = $\text{Var}(X) / a^2$
- Q.E.D.



- **Corollary:** If random variable X has standard deviation σ , then $P(|X-E[X]| \geq k\sigma) \leq 1/k^2$
 - This is consistent with the notion of standard deviation (σ) or variance (σ^2) measuring the spread of the PDF around the mean (center of mass)

Chebyshev's Inequality



Chebyshev

- Pafnuty Chebyshev
 - Founding father of Russian mathematics
 - Students: Lyapunov, Markov
 - First person to think systematically in terms of random variables and their moments and expectations



Markov

- Andrey Markov

- Russian mathematician best known for his work on stochastic processes
- Advisor: Chebyshev
- Students: Voronoy
- One year after doctoral defense, appointed extraordinary professor
- He figured out that he could use chains to model the alliteration of vowels and consonants in Russian literature



А. А. Марков (1886).

Jensen's Inequality

- **Theorem:** Let X be any random variable; $f(\cdot)$ be any **convex** function.
Then, $E[f(X)] \geq f(E[X])$

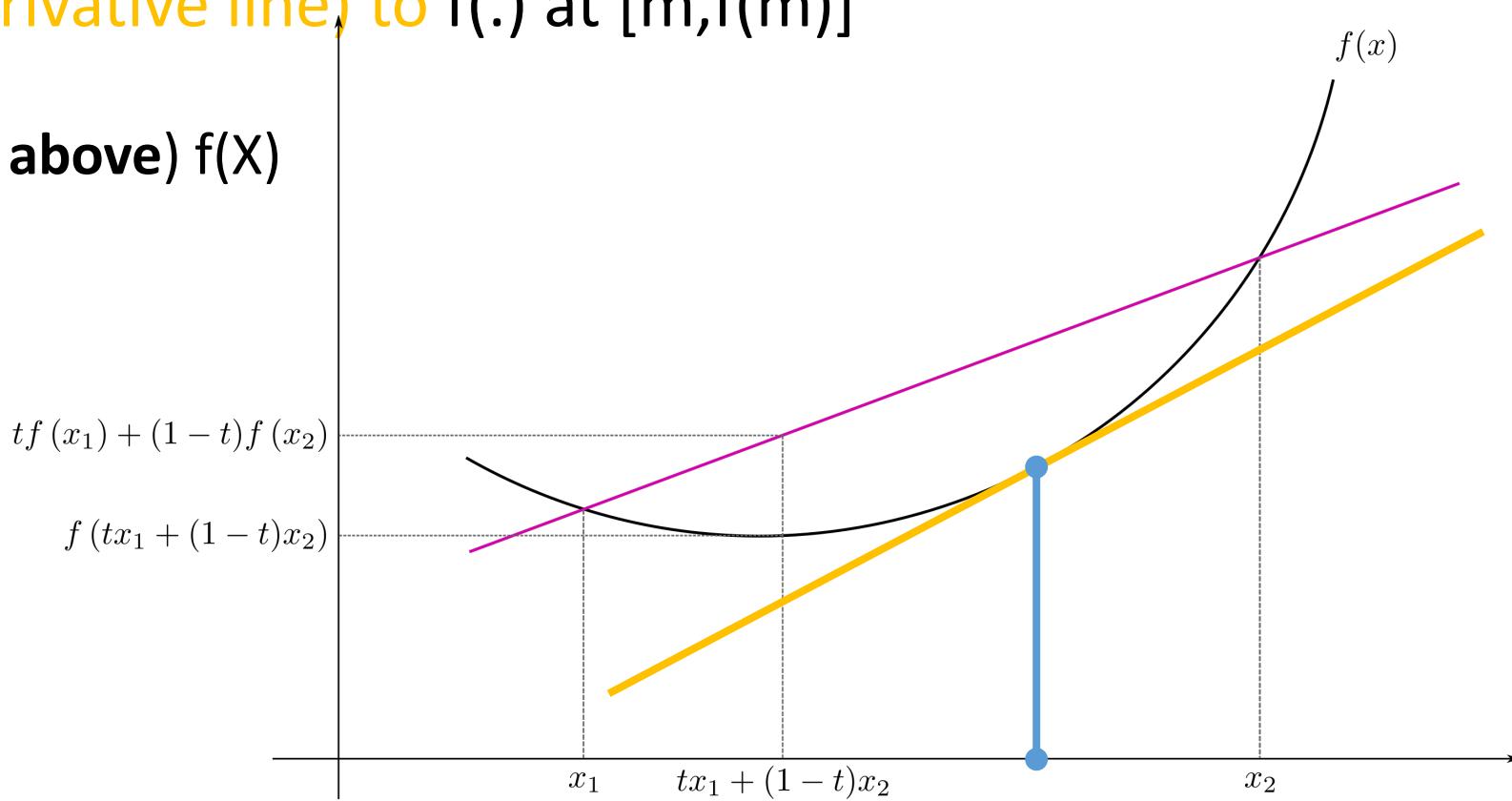
- Proof:

- Let $m := E[X]$, can be anywhere on real line
- Consider a **tangent (subderivative line)** to $f(\cdot)$ at $[m, f(m)]$

- This line is, say, $Y = aX+b$,
which lies at/below (**never above**) $f(X)$
- Then, $f(m) = am+b$

- Then,
$$E[f(X)] \geq E[aX+b]$$
$$= aE[X] + b$$
$$= f(E[X])$$

A real-valued function is called **convex** if the **line segment** between any two points on the graph of the function lies **above/never-below** the graph between the two points.



Jensen's Inequality

- **Corollary:** Let X be any random variable; $g(\cdot)$ be any **concave** function. Then, $E[g(X)] \leq g(E[X])$

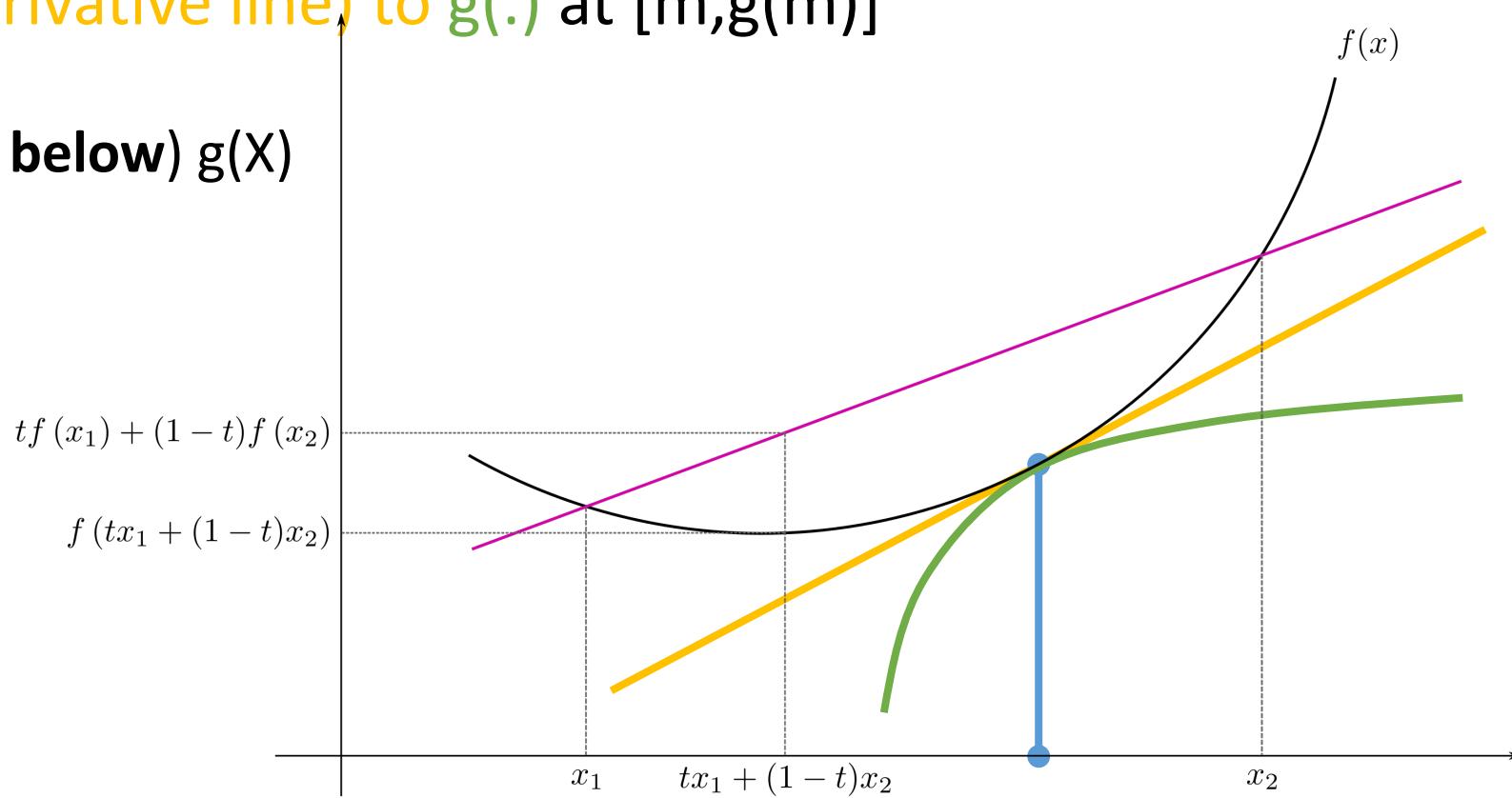
- Proof:

- Let $m := E[X]$, can be anywhere on real line
- Consider a **tangent (subderivative line)** to $g(\cdot)$ at $[m, g(m)]$

- This line is, say, $Y = aX+b$, which lies at/above (never below) $g(X)$
- Then, $g(m) = am+b$

- Then,
$$E[g(X)] \leq E[aX+b]$$
$$= aE[X] + b$$
$$= g(E[X])$$

A real-valued function is called **concave** if the **line segment** between any two points on the graph of the function lies **below/never-above** the graph between the two points.



Jensen

- Johan Jensen
 - Danish mathematician and engineer
 - President of the Danish Mathematical Society from 1892 to 1903
 - Never held any academic position
 - Engineer for Copenhagen Telephone Company
 - Became head of its technical department
 - Learned advanced math topics by himself
 - All his mathematics research was carried out in his spare time



Minimizer of Expected Absolute Deviation

- **Theorem:** $E[|X - c|]$ is minimum when $c = \text{Median}(X)$
- Case 1: Let $c \leq m := \text{Median}(X)$
 - $E[|X - c|] = \int_{-\infty}^c (c - x)P(x)dx + \int_c^{\infty} (x - c)P(x)dx$ (say, **A + B**)
 - $A = \int_{-\infty}^m (c - x)P(x)dx - \int_c^m (c - x)P(x)dx$ (say, **A1 - A2**)
 - $B = \int_c^m (x - c)P(x)dx + \int_m^{\infty} (x - c)P(x)dx$ (say, **B1 + B2**)
 - Now, $B1 - A2 = 2 \int_c^m (x - c)P(x)dx \geq 0$
 - $A1 = \int_{-\infty}^m (c - m)P(x)dx + \int_{-\infty}^m (m - x)P(x)dx$ (say, **A11 + A12**)
 - $B2 = \int_m^{\infty} (x - m)P(x)dx + \int_m^{\infty} (m - c)P(x)dx$ (say, **B21 + B22**)
 - Now, $A11 + B22 = -(m - c)(1 - P(x \geq m)) + (m - c)P(x \geq m) = (m - c)(2P(x \geq m) - 1) \geq 0$
 - Now, $A12 + B21 = E[|X - m|]$
 - So, $\mathbf{A+B} = E[|X - m|] + (m - c)(2P(x \geq m) - 1) + 2 \int_c^m (x - c)P(x)dx$
 - Value of c minimizing **A+B** is $c = m$

Minimizer of Expected Absolute Deviation

- **Theorem:** $E[|X - c|]$ is minimum when $c = \text{Median}(X)$
- Case 2: Let $m := \text{Median}(X) \leq c$
 - $E[|X - c|] = \int_{-\infty}^c (c - x)P(x)dx + \int_c^{\infty} (x - c)P(x)dx$ (say, **A + B**)
 - $A = \int_{-\infty}^m (c - x)P(x)dx + \int_m^c (c - x)P(x)dx$ (say, **A₁ + A₂**)
 - $B = - \int_m^c (x - c)P(x)dx + \int_c^{\infty} (x - c)P(x)dx$ (say, **-B₁ + B₂**)
 - Now, $A_2 - B_1 = 2 \int_m^c (c - x)P(x)dx \geq 0$
 - $A_1 = \int_{-\infty}^m (c - m)P(x)dx + \int_{-\infty}^m (m - x)P(x)dx$ (say, **A₁₁ + A₁₂**)
 - $B_2 = \int_m^{\infty} (x - m)P(x)dx + \int_m^{\infty} (m - c)P(x)dx$ (say, **B₂₁ + B₂₂**)
 - Now, $A_{11} + B_{22} = (c - m) P(x \leq m) - (c - m) (1 - P(x \leq m)) = (c - m) (2P(x \leq m) - 1) \geq 0$
 - Now, $A_{12} + B_{21} = E[|X - m|]$
 - So, $\mathbf{A+B} = E[|X - m|] + (c - m) (2P(x \leq m) - 1) + 2 \int_m^c (c - x)P(x)dx$
 - Value of c minimizing **A+B** is $c = m$

Mean, Median, Standard Deviation

- **Theorem:**

Mean(X) and Median(X) are within a distance of $SD(X)$ of each other

- **Proof:**

- Distance between mean and median

$$= |E[X] - \text{Median}(X)|$$

$$= |E[X - \text{Median}(X)]|$$

This is $|E[.]|$, where $|.|$ is a convex function. Apply Jensen's inequality.

$$\leq E[|X - \text{Median}(X)|]$$

$\leq E[|X - E[X]|]$ (because Median(X) minimizes expected absolute deviation)

$$= E[\text{Sqrt}\{ (X - E[X])^2 \}]$$

This is $E[\text{Sqrt}(.)]$, where $\text{Sqrt}(.)$ is a concave function. Apply Jensen's inequality.

$$\leq \text{Sqrt}\{ E[(X - E[X])^2] \}$$

$$= \text{Sqrt}\{ \text{Var}(X) \} = SD(X)$$

Law of Large Numbers

- This justifies why the expectation is motivated as an average over a large number of random experiments (“long-term average”)
- Let random variables $X_1, \dots, X_i, \dots, X_n$ be ‘n’ **independent and identically distributed (i.i.d.)**, each with mean $\mu=E[X_i]$ and finite variance $v=Var(X_i)$
- Let the **average**, over ‘n’ experiments, be modeled by a random variable $\bar{X} := (X_1 + \dots + X_n) / n$
- Then, the **expected average** $E[\bar{X}] = \mu$, by the linearity of expectation
- But, in specific runs, how close is \bar{X} to the expectation μ ?
- So, we analyze the spread of \bar{X} around μ
- $Var(\bar{X}) = Var(X_1/n) + \dots + Var(X_n/n) = n(v/n^2) = v/n$

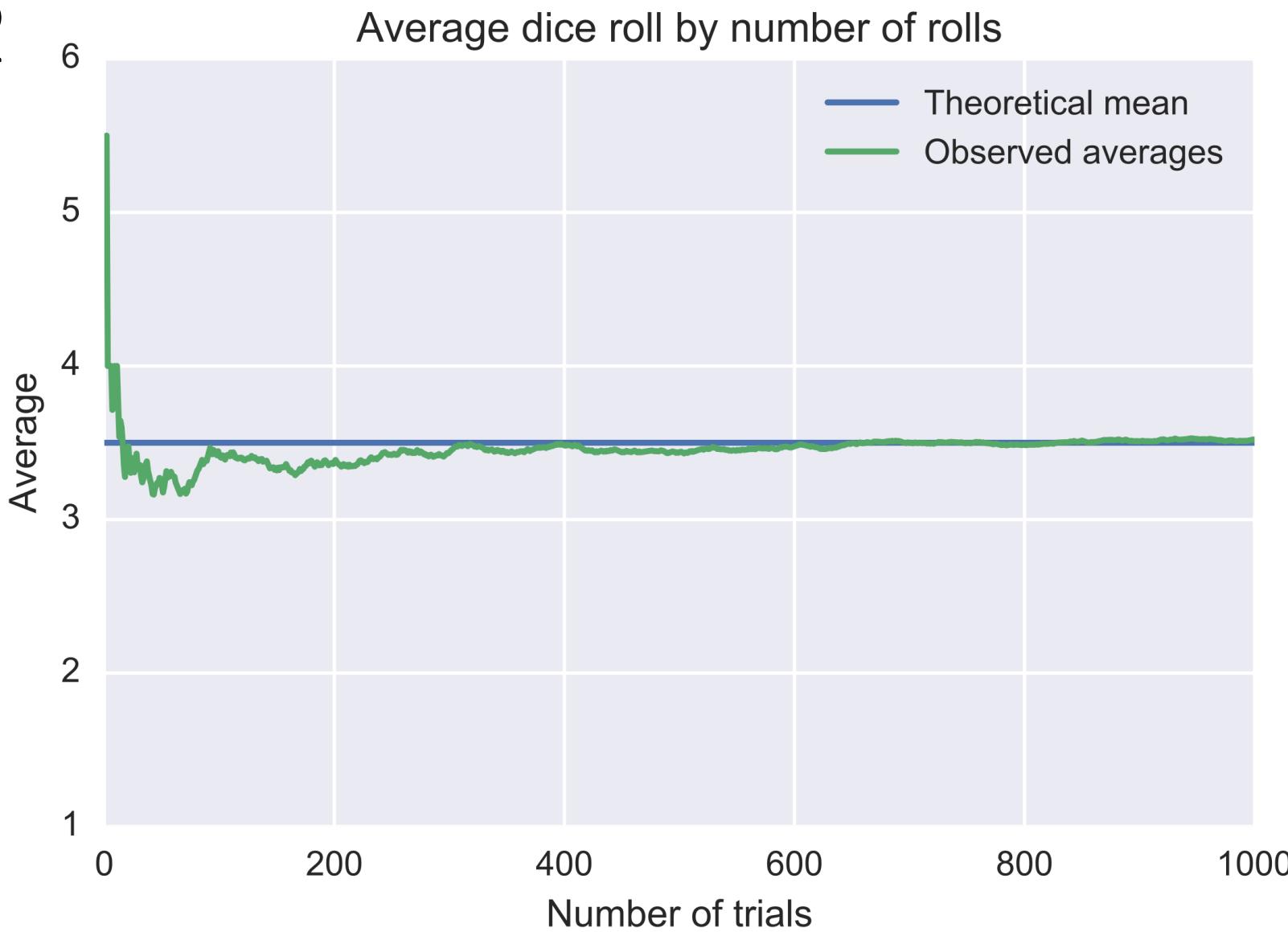
Law of Large Numbers

- This justifies why the expectation is motivated as an average over a large number of random experiments
- **Law of large numbers:** For all $\varepsilon > 0$, as $n \rightarrow \infty$, $P(|\bar{X} - \mu| \geq \varepsilon) \rightarrow 0$
- Proof: Using Chebyshev's inequality,
$$\begin{aligned} P(|\bar{X} - \mu| \geq \varepsilon) \\ \leq \text{Var}(\bar{X}) / \varepsilon^2 \\ = v / (n\varepsilon^2) \\ \rightarrow 0, \text{ as } n \rightarrow \infty \end{aligned}$$
- Thus, as the average \bar{X} uses data from more number of experiments 'n', the event of " \bar{X} being farther from μ than ε " has a probability that tends to 0

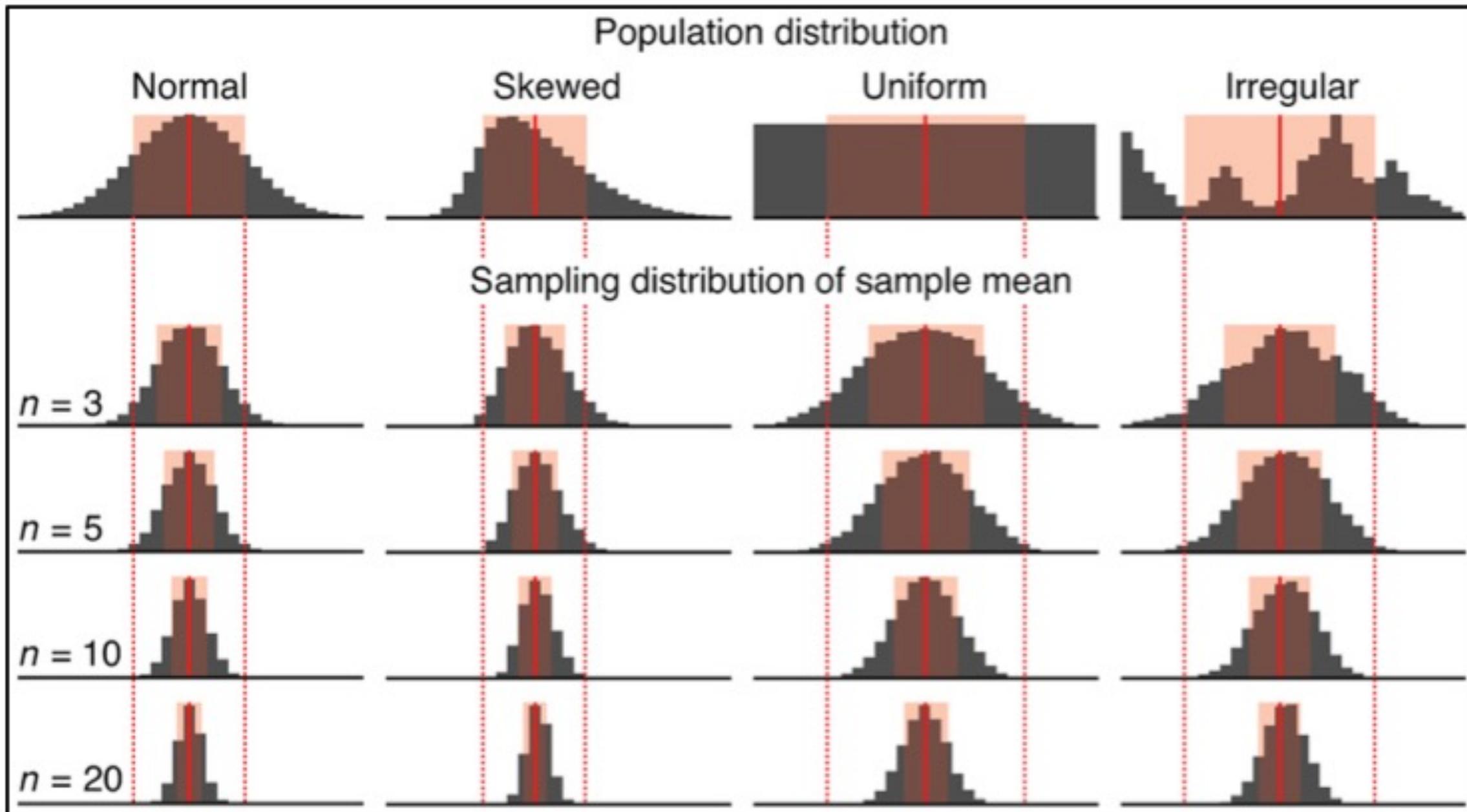
Law of Large Numbers

- Example

- This also gives us a way to compute an “estimate” of the **expectation μ of a random variable X** from “observations”/data
 - What is the estimate ?
 - \bar{X}



Law of Large Numbers

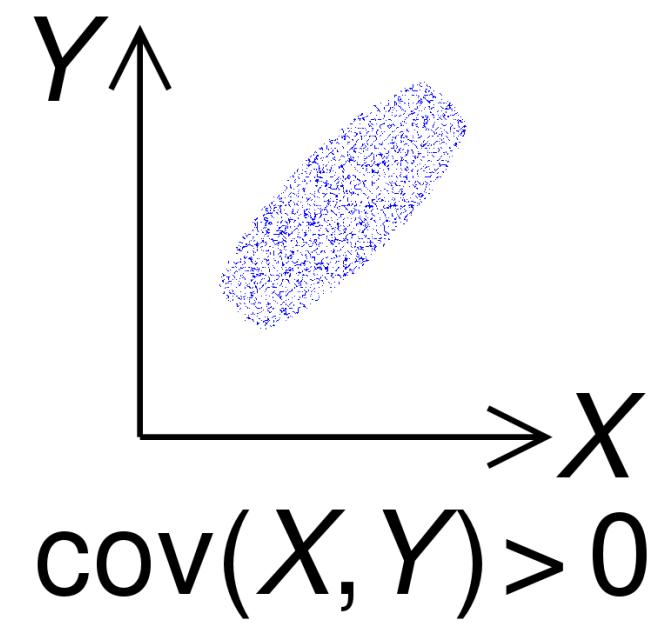
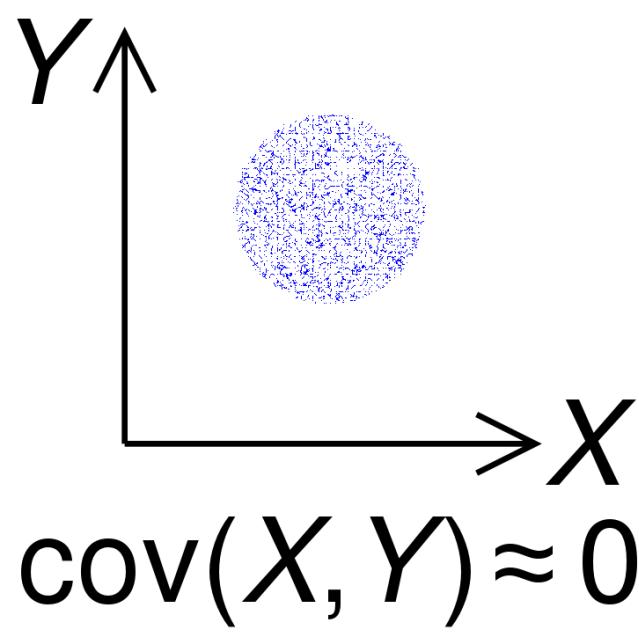
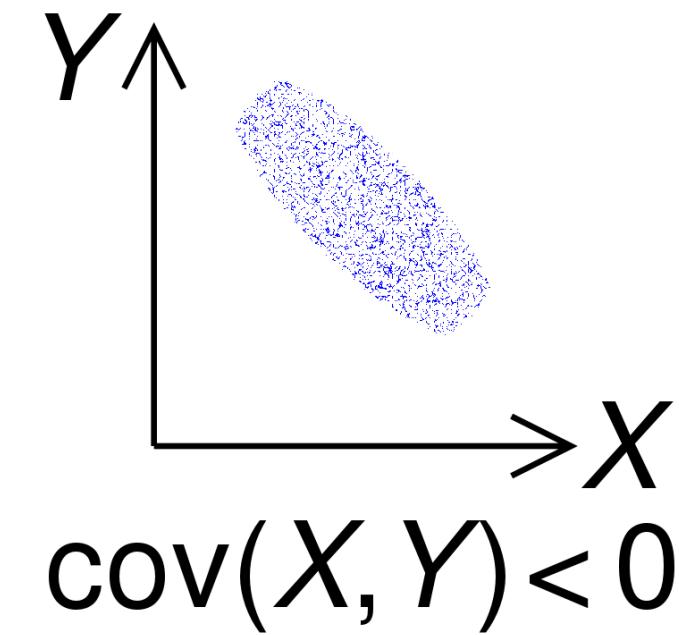


Covariance

- For random variables X and Y , consider the joint PMF/PDF $P(X,Y)$
- Covariance: A measure of how the values taken by X and Y vary together (“co”-“vary”)
- **Definition:** $\text{Cov}(X,Y) := E[(X - E[X])(Y - E[Y])]$
 - Interpretation:
 - Define $U(X) := X - E[X]$ and $V(Y) := Y - E[Y]$ (Note: U and V have expectation 0)
 - In the joint distribution $P(U,V)$,
if **larger** (more +ve) values of U typically correspond to **larger** values of V , and
smaller (more -ve) values of U typically correspond to smaller values of V ,
then U and V **co-vary positively**
 - In the joint distribution $P(U,V)$,
if **larger** values of U typically correspond to **smaller** values of V , and ...
then U and V **co-vary negatively**
 - **Property: Symmetry:** $\text{Cov}(X,Y) = \text{Cov}(Y,X)$

Covariance

- Examples



Covariance

- **Property:** $\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$

- Proof:

- $\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y]$
- So, $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2(\text{Cov}(X,Y) - E[X]E[Y]) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$
- Also, when X and Y are independent, then $\text{Cov}(X,Y) = 0$

- **Property:** When $\text{Var}(X)$ and $\text{Var}(Y)$ are finite, and one of them is 0, then $\text{Cov}(X,Y)=0$

- **Property:** When $Y := mX + c$ (with finite m), what is $\text{Cov}(X,Y) ?$

- $$\begin{aligned}\text{Cov}(X,Y) &= E[XY] - E[X]E[Y] \\ &= E[mX^2 + cX] - E[X](m.E[X] + c) \\ &= m.E[X^2] - m(E[X])^2 = m.\text{Var}(X)\end{aligned}$$

- When $\text{Var}(X)>0$, covariance is \propto line-slope 'm', and has same sign as that of m

Covariance

- **Bilinearity of Covariance**

- Let X, X_1, X_2, Y, Y_1, Y_2 be random variables. Let c be a scalar constant.
- **Property:** $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y) = \text{Cov}(Y, X_1 + X_2)$
 - Proof (first part; second part follows from symmetry):

$$\begin{aligned}\text{Cov}(X_1 + X_2, Y) &= E((X_1 + X_2)Y) - E(X_1 + X_2)E(Y) \\ &= E(X_1Y) - E(X_1)E(Y) + E(X_2Y) - E(X_2)E(Y) \\ &= \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)\end{aligned}$$

- **Property:** $\text{Cov}(aX, Y) = a \cdot \text{Cov}(X, Y) = \text{Cov}(X, aY)$

- Proof (first part):

- $\text{Cov}(aX, Y)$

$$= E[aXY] - E[aX]E[Y]$$

$$= a(E[XY] - E[X]E[Y])$$

$$= a \text{Cov}(X, Y)$$

Standardized Random Variable

- **Definition:**

If X is a random variable, then its **standardized form** is given by $X^* := (X - E[X]) / SD(X)$, where $SD(\cdot)$ gives the standard deviation

- **Property:** $E[X^*] = 0$, $\text{Var}(X^*) = 1$

- Proof:

$$E(X^*) = \frac{E(X) - E(X)}{\sigma_X} = 0$$

$$\text{Var}(X^*) = \text{Var}\left(\frac{X - E(X)}{\sigma_X}\right) = \frac{\text{Var}(X)}{\sigma_X^2} = \frac{\text{Var}(X)}{\text{Var}(X)} = 1$$

- X^* is unit-less

- X^* is obtained by:

- First shifting/translating X to make mean 0, and
- Then scaling the shifted variable to make variance 1

Correlation

- For covariance, the magnitude isn't easy to interpret (unlike its sign)
- **Correlation:** A measure of how the values taken by X and Y vary together ("co"-“relate”) obtained by rescaling covariance
 - Pearson's correlation coefficient
 - Assuming X and Y are linearly related, correlation magnitude shows the strength of the (functional/deterministic) relationship between X and Y
- Let 'SD' = standard deviation
- **Definition:** $\text{Cor}(X,Y) := E \left[\left(\frac{X - E[X]}{\text{SD}(X)} \right) \left(\frac{Y - E[Y]}{\text{SD}(Y)} \right) \right] = \text{Cov}(X, Y) / (\text{SD}(X)\text{SD}(Y))$
 - Thus, $\text{Cor}(X,Y) = E[X^*Y^*]$, where X^* and Y^* are the standardized variables
 $= E[X^*Y^*] - E[X^*]E[Y^*]$
 $= \text{Cov}(X^*, Y^*)$

Correlation

- **Property:** $-1 \leq \text{Cor}(X,Y) \leq 1$

- Proof:

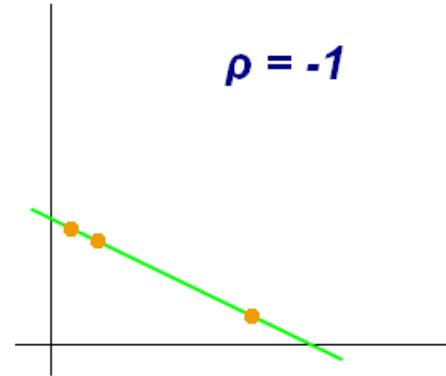
- First inequality

$$0 \leq E[(X^* + Y^*)^2]$$

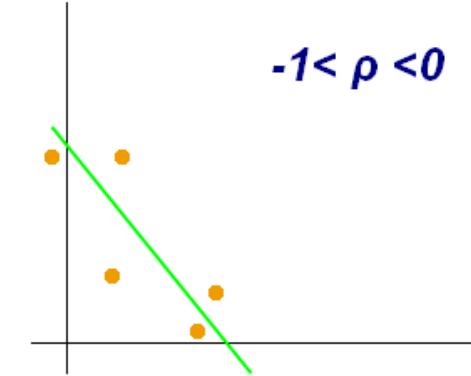
$$= E[(X^*)^2] + E[(Y^*)^2] + 2E[X^*Y^*]$$

$$= 2(1 + \text{Cor}(X,Y))$$

- So, $-1 \leq \text{Cor}(X,Y)$



$$\rho = -1$$



$$-1 < \rho < 0$$

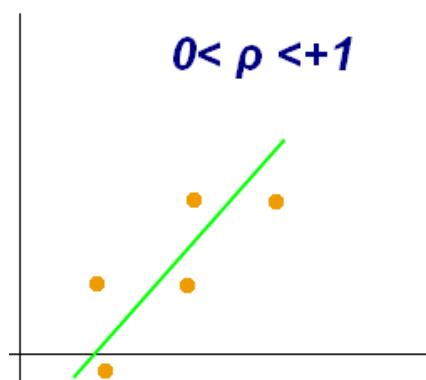
- Second inequality

$$0 \leq E[(X^* - Y^*)^2]$$

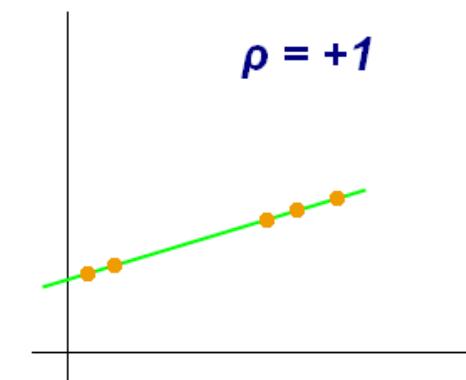
$$= E[(X^*)^2] + E[(Y^*)^2] - 2E[X^*Y^*]$$

$$= 2(1 - \text{Cor}(X,Y))$$

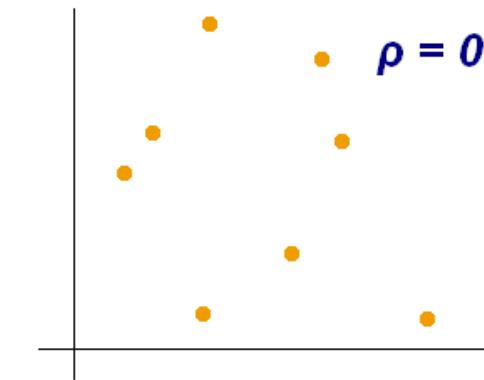
- So, $\text{Cor}(X,Y) \leq 1$



$$0 < \rho < +1$$



$$\rho = +1$$



$$\rho = 0$$

Correlation

- **Property:** If X and Y are linearly related, i.e., $Y = mX + c$, and are non-constant (i.e., $SD(X) > 0$ and $SD(Y) > 0$), then $|\text{Cor}(X,Y)| = 1$

- Proof:

- When $Y = mX + c$, then $SD(Y) = |m| SD(X)$

- $\text{Cor}(X,Y)$

$$= \text{Cov}(X,Y) / (SD(X) SD(Y))$$

$$= m\text{Var}(X) / (SD(X) |m| SD(X))$$

$$= \pm 1$$

$$= \text{sign of the slope } m$$


Correlation

- **Property:** If $|\text{Cor}(X,Y)| = 1$, then X and Y are linearly related
 - Proof:
 - If $\text{Cor}(X,Y) = 1$, then $E[(X^* - Y^*)^2] = 2(1 - \text{Cor}(X,Y)) = 0$
 - For discrete X,Y: this must imply $X^* = Y^*$ for all (x',y') where $P(X=x',Y=y') > 0$
 - Else the summation underlying the expectation cannot be zero
 - For continuous X,Y: this must imply $X^* = Y^*$ for all measures (dx',dy') where $P(dx',dy') > 0$
 - X^* and Y^* can be unequal only on a countable set of isolated points where $P(dx',dy') > 0$
 - Else the integral underlying the expectation cannot be zero
 - If $\text{Cor}(X,Y) = (-1)$, then $E[(X^* + Y^*)^2] = 2(1 + \text{Cor}(X,Y)) = 0$
 - For discrete X,Y: this must imply $X^* = -Y^*$ for all (x',y') where $P(X=x',Y=y') > 0$
 - For continuous X,Y: this must imply $X^* = -Y^*$ for all measures (dx',dy') where $P(dx',dy') > 0$
 - Inequality can hold only on a countable set of isolated points where $P(dx',dy') > 0$
 - If $X^* = \pm Y^*$, then Y must be of the form $mX+c$
 - For continuous X,Y: this must imply $X^* = Y^*$ for all measures (dx',dy') where $P(dx',dy') > 0$
 - X^* and Y^* can be unequal only on a countable set of isolated points where $P(dx',dy') > 0$
 - Else the integral underlying the expectation cannot be zero
 - If $\text{Cor}(X,Y) = (-1)$, then $E[(X^* + Y^*)^2] = 2(1 + \text{Cor}(X,Y)) = 0$
 - For discrete X,Y: this must imply $X^* = -Y^*$ for all (x',y') where $P(X=x',Y=y') > 0$

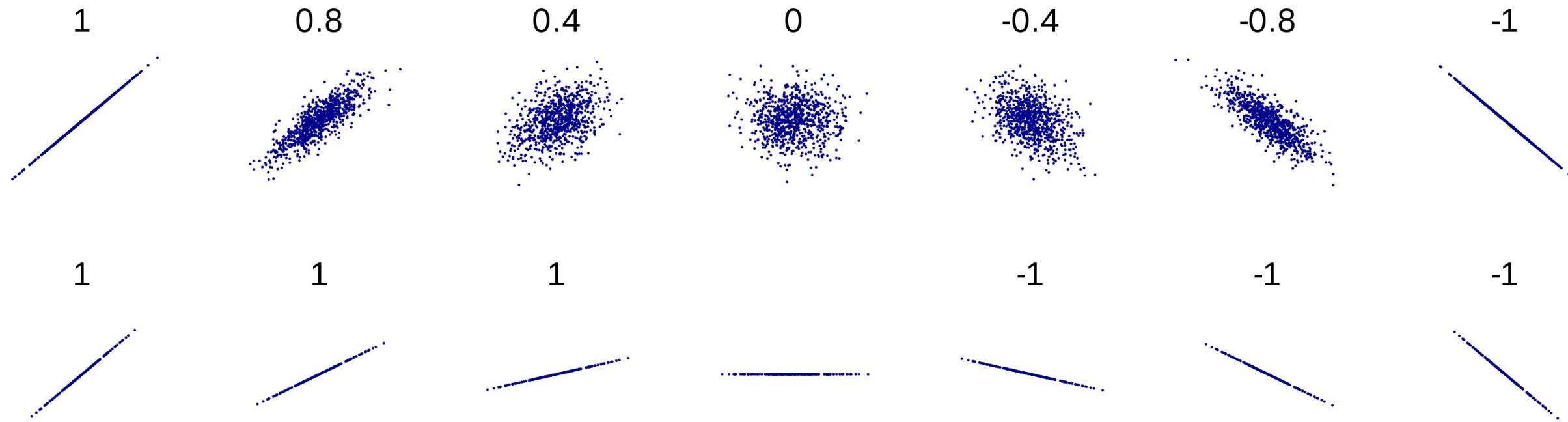
Correlation

- If $|\text{Cor}(X,Y)|=1$ (or $Y=mX+c$), then how to find the equation of the line from data $\{(x_i, y_i) : i=1, \dots, n\}$?
 - By the way: line must pass through $(E[X], E[Y])$
 - Because, when $X=E[X]$, value of Y must be $mE[X]+c$, but that also equals $E[Y]$
 - We proved that: if $Y=mX+c$, then $|\text{Cor}(X,Y)|=1$ and $Y^* = \pm X^* = \text{Cor}(X,Y) X^*$
 - So, $(Y - E[Y]) / SD(Y) = \text{Cor}(X,Y) (X - E[X]) / SD(X)$
 - So, $Y = E[Y] + SD(Y) \text{Cor}(X,Y) (X - E[X]) / SD(X)$
 - So, $Y = E[Y] + \text{Cov}(X,Y) (X - E[X]) / \text{Var}(X)$
 - This gives the equation of the line with:
 - Slope $m := \text{Cov}(X,Y) / \text{Var}(X)$
 - Intercept $c := E[Y] - \text{Cov}(X,Y) E[X] / \text{Var}(X)$



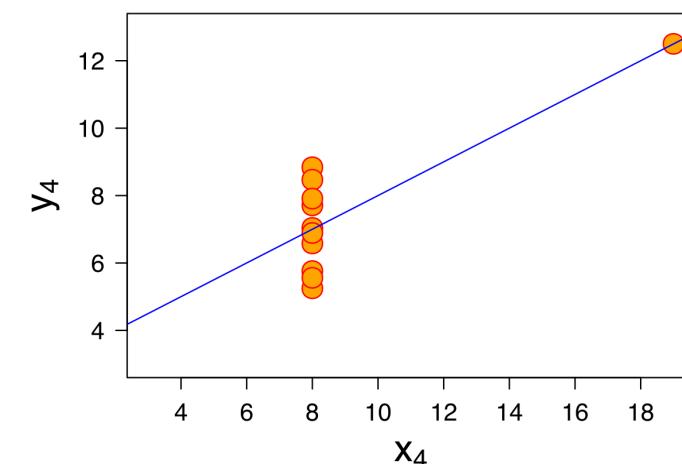
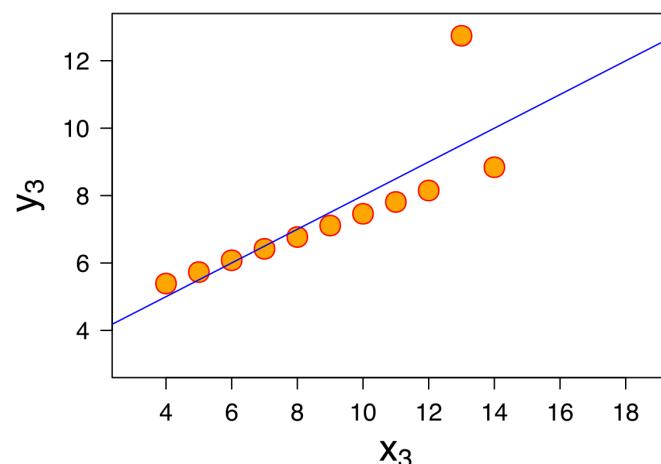
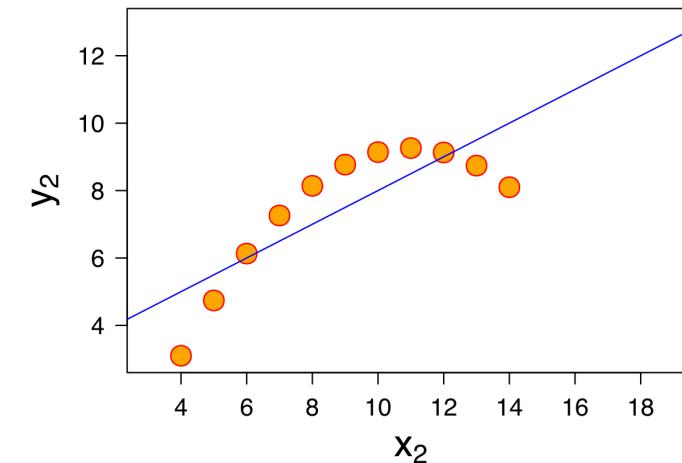
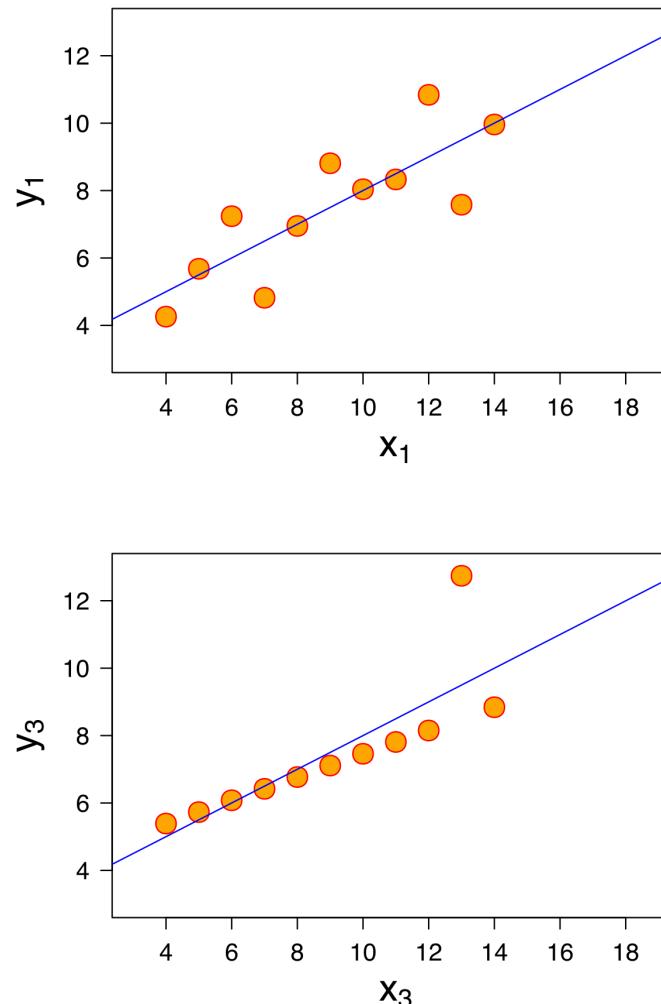
Correlation

- Examples



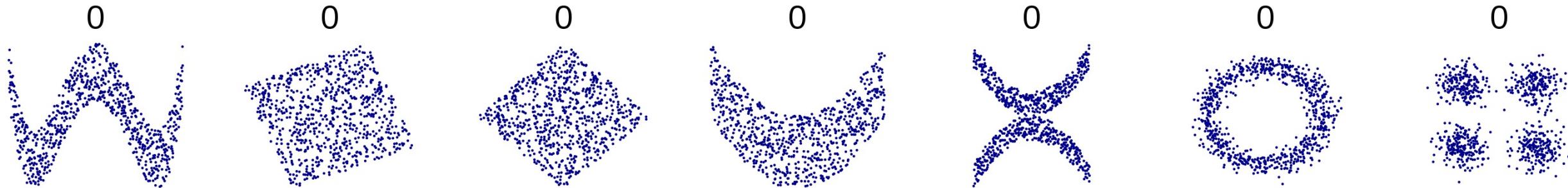
Correlation

- Four sets of data with the same correlation of 0.816
 - Blue line indicates the line passing through $(E[X], E[Y])$ with slope = 0.816 (more on this when we study estimation)
 - So, correlation = 0.816 doesn't always mean that data lies along a line of slope 0.816
 - This indicates the likely misinterpretation of correlation when variables underlying data aren't linearly related



Correlation

- Zero correlation doesn't imply independence



- We showed that independence implies zero covariance/correlation, but the converse isn't always true
- Example: Let X be uniformly distributed within $[-1,+1]$. Let $Y := X^2$.
 - $\text{Cov}(X,X^2) = E[X \cdot X^2] - E[X]E[X^2] = E[X^3] - 0 \cdot E[X^2] = 0$
 - Thus, $\text{Cov}(X,Y) = 0 = \text{Cor}(X,Y)$ even though Y is a deterministic function of X

Correlation

- **Non-zero correlation doesn't imply causation**

- <https://hbr.org/2015/06/beware-spurious-correlations>
- <https://science.sciencemag.org/content/348/6238/980.2>
- <http://www.tylervigen.com/spurious-correlations>

BOOKS ET AL. | STATISTICS

Spurious Correlations

+ See all authors and affiliations

Science 29 May 2015:
Vol. 348, Issue 6238, pp. 980
DOI: 10.1126/science.aac5518

Article

Figures & Data

Info & Metrics

eLetters

 PDF

Summary

Presented as a series of graphs prepared from real data sets, *Spurious Correlations* serves as a hilarious reminder that correlation most certainly does not equal causation.

[View Full Text](#)

Correlation

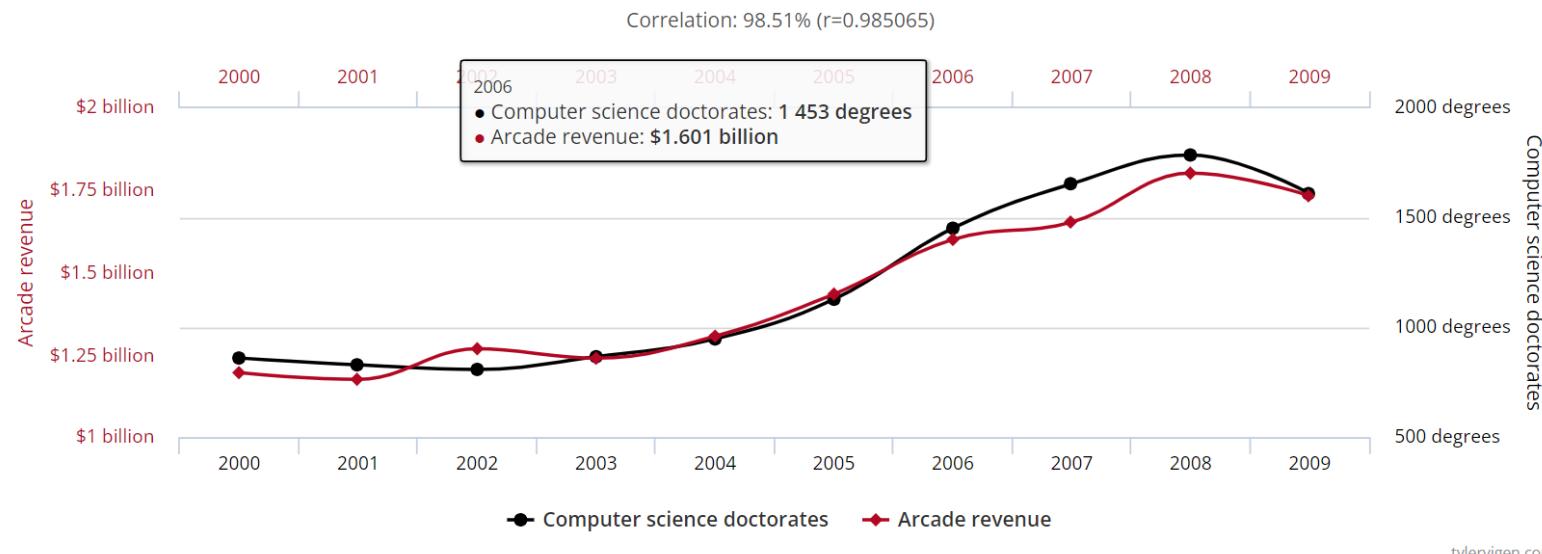
- Non-zero correlation does

- <https://hbr.org/2015/06/beyond-correlation>
- <https://science.sciencemag.org/>
- <http://www.tylervigen.com>

Total revenue generated by arcades

correlates with

Computer science doctorates awarded in the US

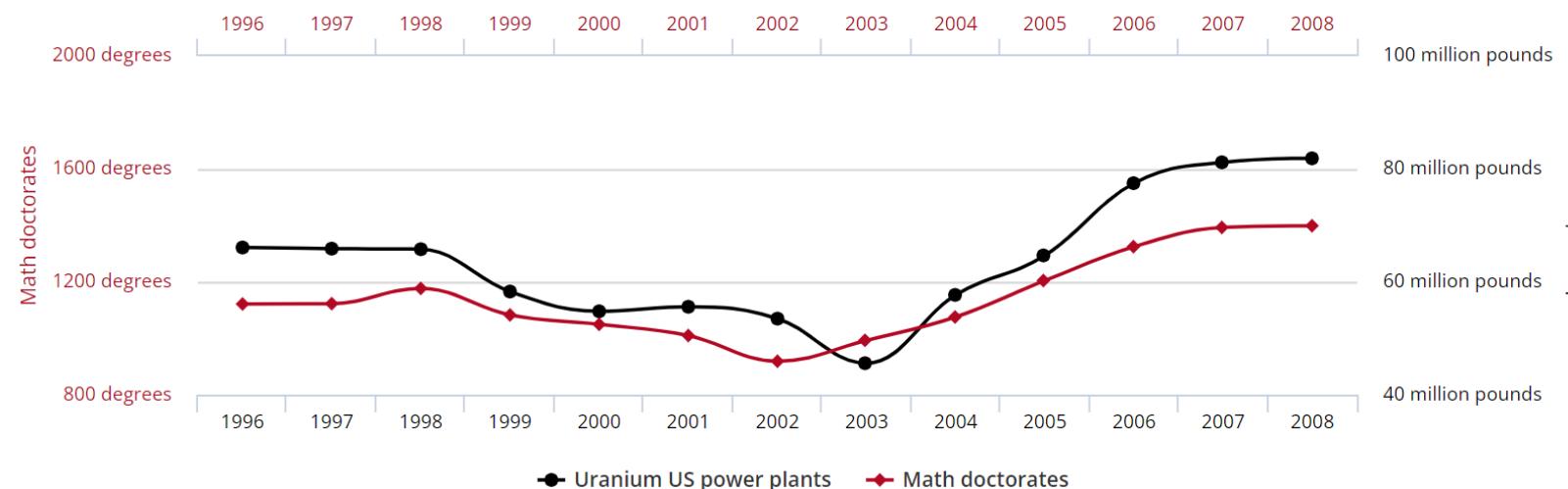


Math doctorates awarded

correlates with

Uranium stored at US nuclear power plants

Correlation: 95.23% ($r=0.952257$)



Correlation

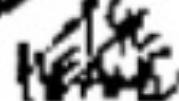
- Non-zero correlation doesn't imply causation

THE FAMILY CIRCUS



8-5

© 1988 Bill Amend, Inc.
McCourt Syndicate, Inc.



"I wish they didn't turn on that seatbelt sign so much! Every time they do, it gets bumpy."

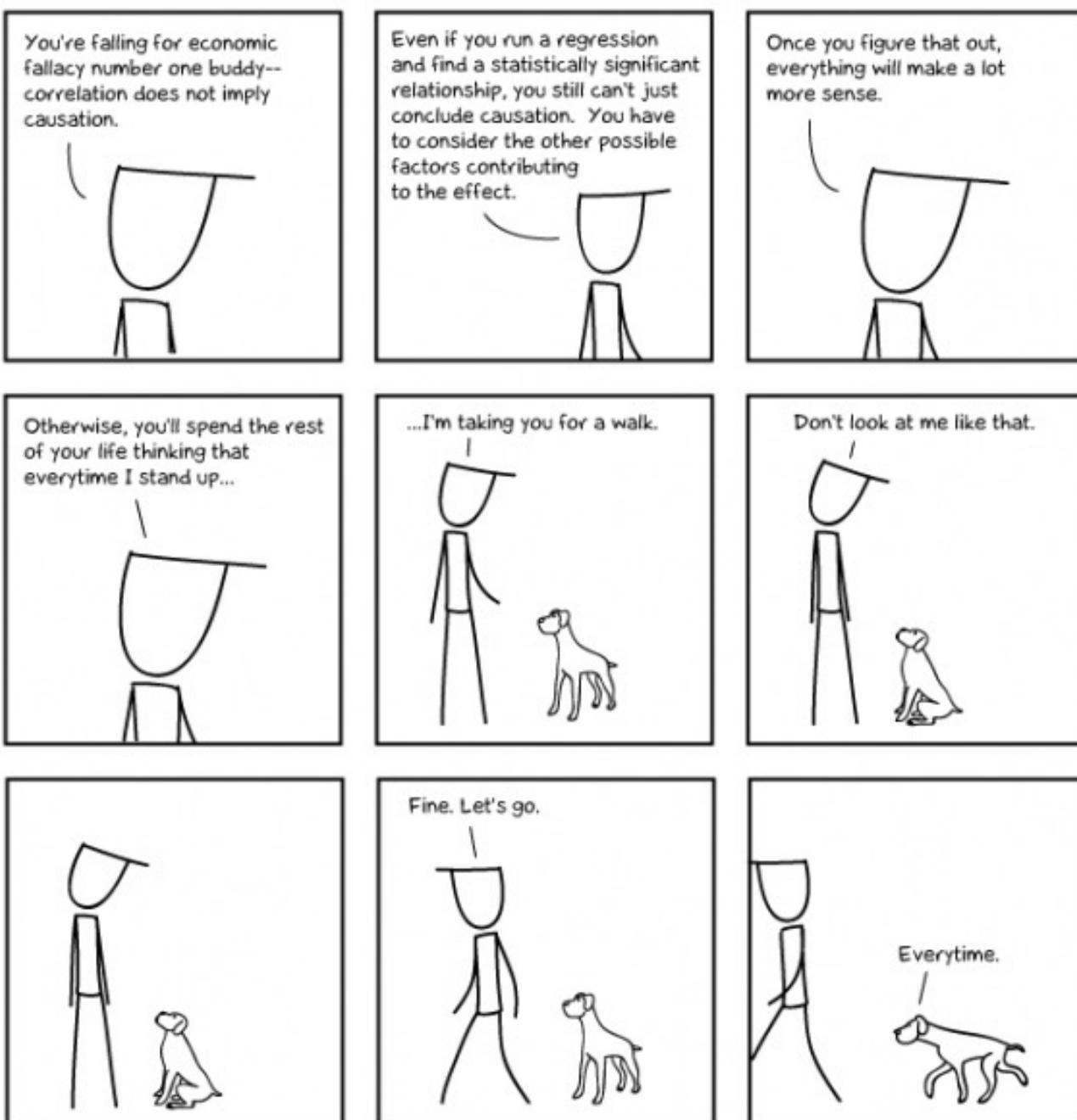
Correlation

- Non-zero correlation doesn't imply causation



Correlation

- Non-zero correlation doesn't imply causation



Doghouse Diaries
"Better than a poke in the eye with a sharp stick."