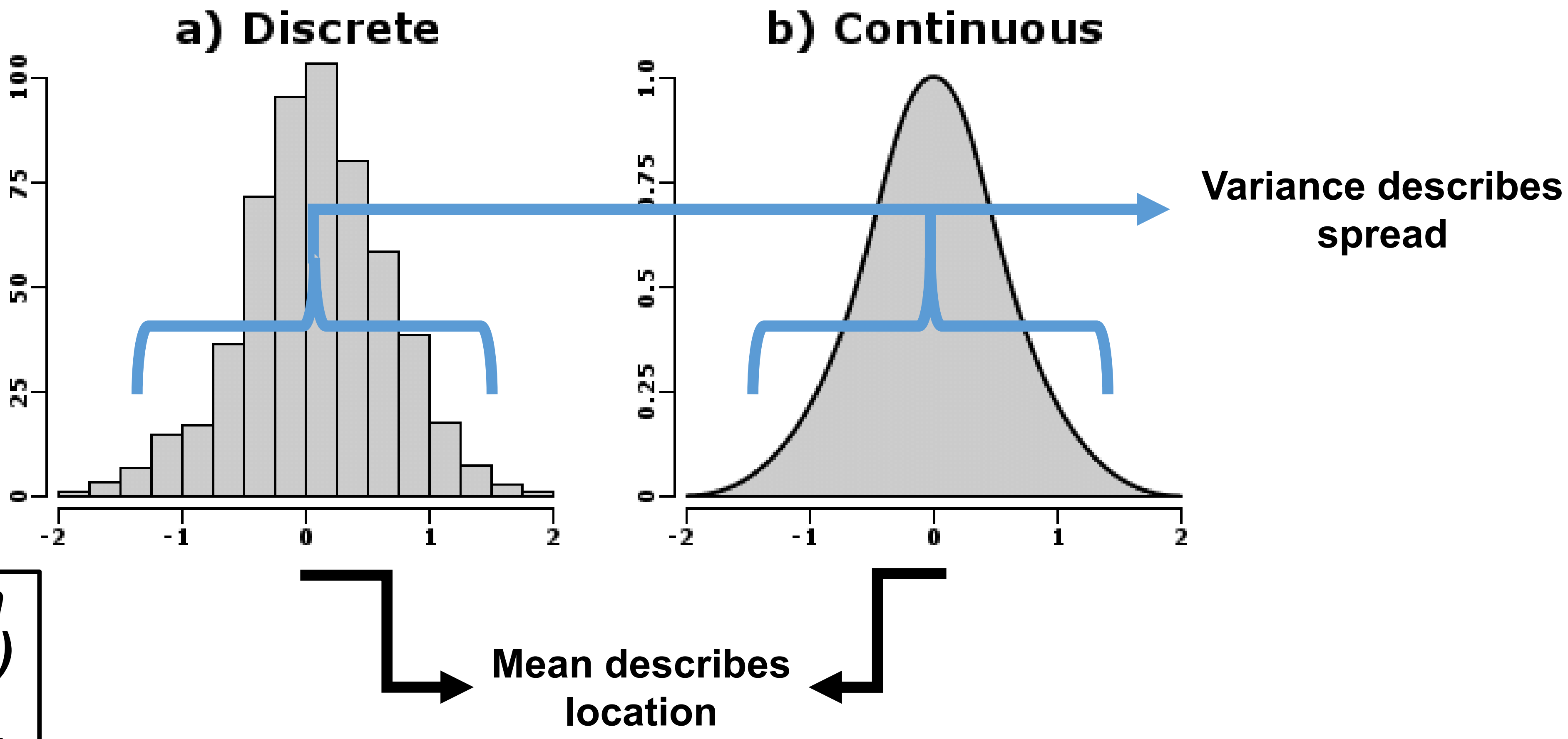
- Expected Value
- Variance, Covariance, Correlation
- Dependence of Random Variables

- Expected Value
- Variance, Covariance, Correlation
- Dependence of Random Variables

Moments of Random Variables

Properties of a RV are characterized by its distribution / PMF / PDF



We will focus on (mean / variance)

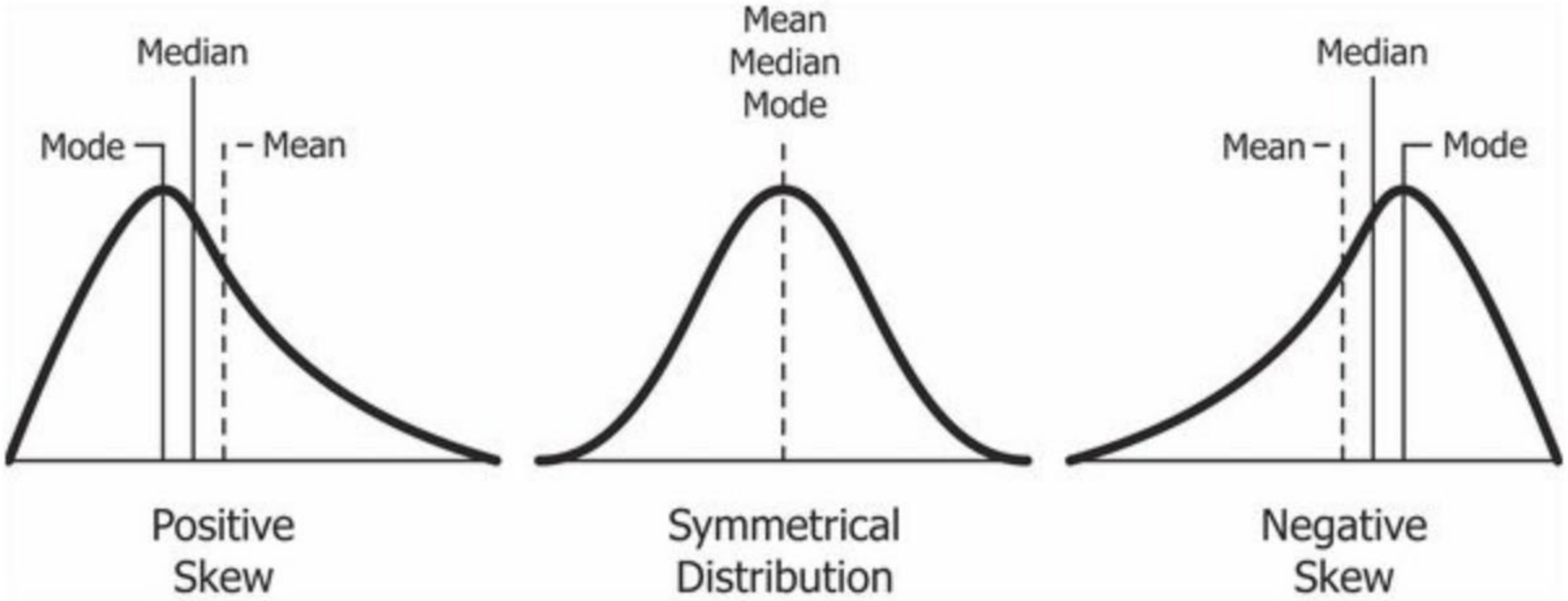
But others exist…

Moments characterize properties of the distribution "shape"

Moments of Random Variables

Higher-order moments characterize other aspects of distribution shape

Example *Skew describes asymmetry of the PMF / PDF*



Additional moments (i.e. kurtosis) are typically less common in data science

Expected Value

Definition *The expectation of a discrete RV X, denoted by $\mathbf{E}[X]$, is:*

$$\mathbf{E}[X] = \sum_x x\, p(X = x)$$

Summation over all values in domain of X

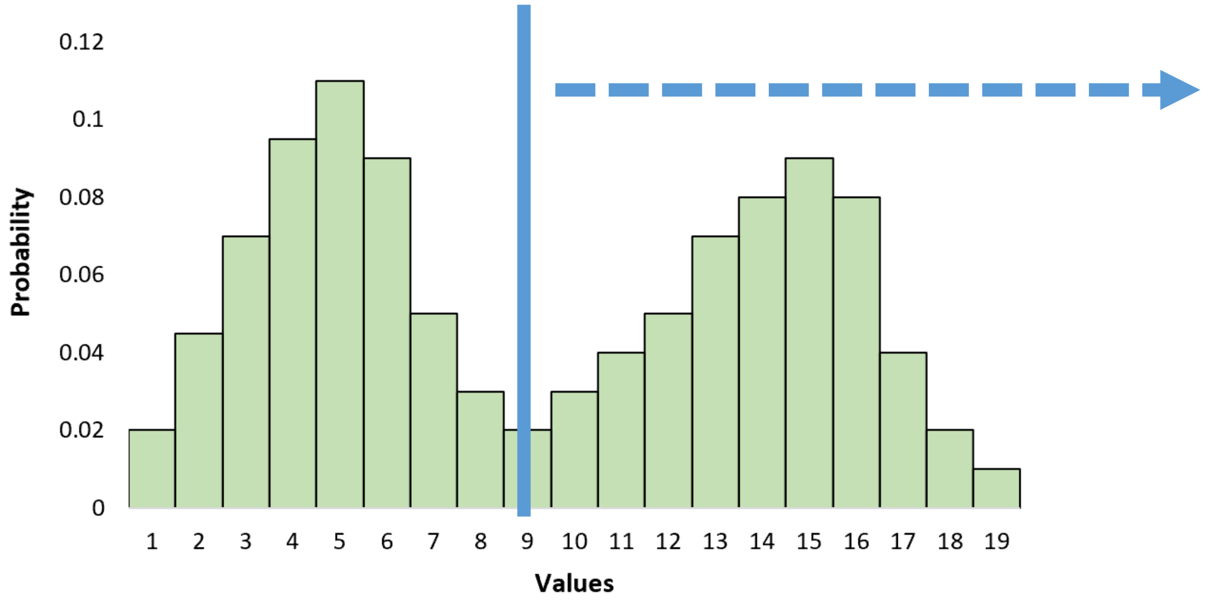
Example Let X be the sum of two fair dice, then:

$$\mathbf{E}[X] = \frac{1}{36} \cdot 2 + \frac{1}{36} \cdot 3 + \ldots + \frac{1}{36} \cdot 12 = 7$$

For ordered dice there are 36 terms in the sum, not 11

- *Also known as the average or mean value in other contexts*
- *Weighted average: each outcome weighted by probability of occurring*
- *Simple average in the case of a uniform distribution*

Expected Value



Expected value is not always a high probability event…

…in fact, it may not even be feasible…

Example Let X be the result of a fair six-sided die, then:

$$\mathbf{E}[X] = \frac{1}{6} \cdot (1 + 2 + 3 + 4 + 5 + 6) = 3.5$$

Can't actually roll 3.5

Expected Value

Theorem (Linearity of Expectations) *For any finite collection of discrete RVs* $X_1, X_2, \ldots, X_N$ *with finite expectations,*

$$\mathbf{E}\left[\sum_{i=1}^{N} X_i\right] = \sum_{i=1}^{N} \mathbf{E}[X_i]$$

E.g. for two RVs X and Y
$\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$

Example Throw two fair six-sided dice. What is the expected sum? Let X and Y be the outcome of the first and second die, respectively. Then,

$$\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y] = 3.5 + 3.5 = 7$$

Expected Value

Proof (Linearity of Expectations)

$$\mathbf{E}[X+Y] = \sum_i \sum_j (i+j) p(X=i, Y=j)$$

By definition of Expectation

$$= \sum_i \sum_j i \cdot p(X=i, Y=j) + \sum_i \sum_j j \cdot p(X=i, Y=j)$$

Sum is linear operator

$$= \sum_i i \sum_j p(X=i, Y=j) + \sum_j j \sum_i p(X=i, Y=j)$$

Law of Total Probability

$$= \sum_i i \cdot p(X=i) + \sum_j j \cdot p(Y=j)$$

By definition of Expectation

$$= \mathbf{E}[X] + \mathbf{E}[Y]$$

Expected Value

Expected value has no effect on a constant

$$\mathbf{E}[c] = c$$

Combined with the linearity of expectations we have that for any random variable X and constant c,

$$\mathbf{E}[cX] = c\mathbf{E}[X]$$

Example Let X and Y be the outcome of two fair six-sided dice, then:

$$\mathbf{E}[2(X+Y)] = 2\mathbf{E}[X] + 2\mathbf{E}[Y]$$

$$= 2 \cdot 3.5 + 2 \cdot 3.5 = 14$$

Expected Value

Definition *The conditional expectation of a discrete RV* X*, given* Y *is:*

$$\mathbf{E}[X \mid Y = y] = \sum_x x\, p(X = x \mid Y = y)$$

Example Roll two standard six-sided dice and let X be the result of the first die and let Y be the sum of both dice, then:

$$\mathbf{E}[X_1 \mid Y = 5] = \sum_{x=1}^{4} x\, p(X_1 = x \mid Y = 5)$$

$$= \sum_{x=1}^{4} x \frac{p(X_1 = x, Y = 5)}{p(Y = 5)} = \sum_{x=1}^{4} x \frac{1/36}{4/36} = \frac{5}{2}$$

Conditional expectation follows properties of expectation (linearity, etc.)

Expected Value

Law of Total Expectation *Let X and Y be discrete RVs with finite expectations, then:*

$$\mathbf{E}[X] = \mathbf{E}_Y[\mathbf{E}_X[X \mid Y]]$$

Proof

$$\mathbf{E}_Y[\mathbf{E}_X[X \mid Y]] = \mathbf{E}_Y\left[\sum_x x \cdot p(x \mid Y)\right]$$

$$= \sum_y \left[\sum_x x \cdot p(x \mid y)\right] \cdot p(y)$$ **(Definition of expectation)**

$$= \sum_y \sum_x x \cdot p(x, y)$$ **(Probability chain rule)**

$$= \sum_x x \sum_y \cdot p(x, y)$$ **(Linearity of expectations)**

$$= \sum_x x \cdot p(x) = \mathbf{E}[X]$$ **(Law of total probability)**

- Expected Value
- Variance, Covariance, Correlation
- Dependence of Random Variables

Moments and Order

We can express different aspects of the distribution from moments of powers of the RV,

$$\mathbf{E}[X^n] = \sum_k X^n p(X = k)$$

- We call these **non-central moments** of order *n*
- *High-order moments* refer to larger powers, typically n>2 or more

Typically, it is more intuitive to first subtract the mean value,

$$\mathbf{E}[(X - \mathbf{E}[X])^n]$$

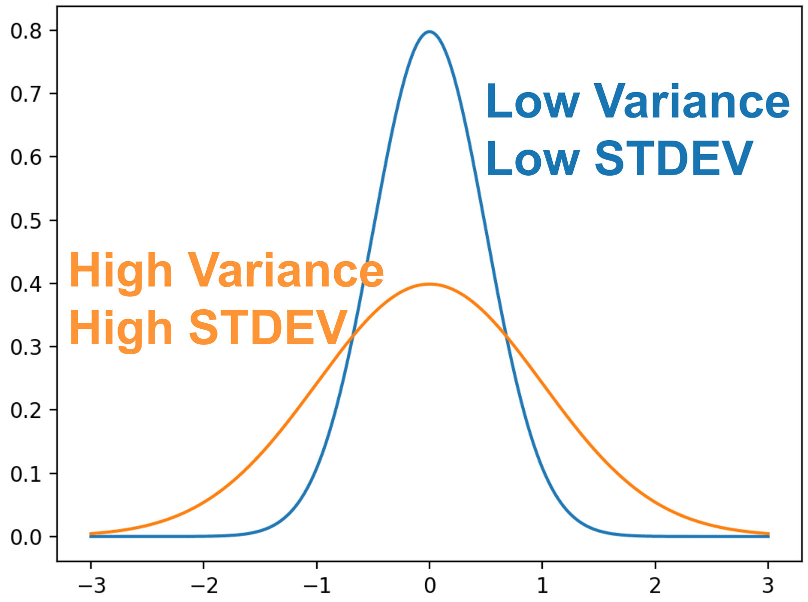
- We call these **central moments**
- The second central moment (n=2) is known as the <u>variance</u>

Variance

Definition The variance of a RV X is defined as,

$$\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$$

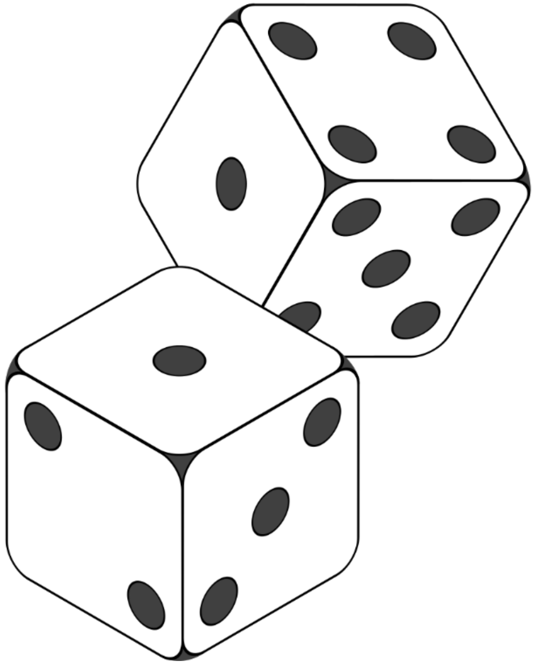
The standard deviation (STDEV) is $\sigma[X] = \sqrt{\mathbf{Var}[X]}$.



- *Describes the "spread" of a distribution*
- *Describes uncertainty of outcome*
- *STDEV is in original units (more intuitive), variance is in units^2*
- *Variance is more mathematically useful than STDEV*

Variance

Example Let X be the result of a fair six-sided die.



The variance is then,

$$\mathrm{Var}(X) = \sum_{i=1}^{6} \frac{1}{6}\left(i - \frac{7}{2}\right)^2$$

$$= \frac{1}{6}\left((-5/2)^2 + (-3/2)^2 + (-1/2)^2 + (1/2)^2 + (3/2)^2 + (5/2)^2\right)$$

$$= \frac{35}{12} \approx 2.92.$$

The STDEV is $\sqrt{\mathrm{Var}(X)} \approx 1.71$, which suggests we should expect outcomes to vary around the mean of 3.5 by +/- 1.71

Variance

Lemma An equivalent form of variance is:

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - (\mathbf{E}[X])^2$$

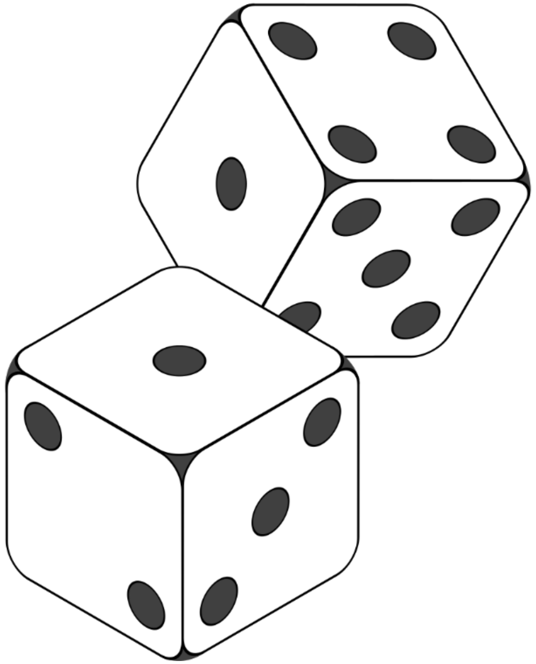
Proof

$$\begin{aligned}
\mathbf{E}[(X - \mathbf{E}[X])^2] &= \mathbf{E}[X^2 - 2X\mathbf{E}[X] + \mathbf{E}[X]^2] && \textbf{(Distributive property)} \\
&= \mathbf{E}[X^2] - 2\mathbf{E}[X]\mathbf{E}[X] + \mathbf{E}[X]^2 && \textbf{(Linearity of expectations)} \\
&= \mathbf{E}[X^2] - 2\mathbf{E}[X]^2 + \mathbf{E}[X]^2 && \textbf{(Algebra)} \\
&= \mathbf{E}[X^2] - \mathbf{E}[X]^2 && \textbf{(Algebra)}
\end{aligned}$$

Variance

Example General form of variance for a fair n-sided die,

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\big(X^2\big) - (\mathrm{E}(X))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} i\right)^2 \\
&= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 \\
&= \frac{n^2-1}{12}.
\end{aligned}$$



Moments of Useful Discrete Distributions

Bernoulli *A.k.a. the **coinflip** distribution on <u>binary</u> RVs* $X \in \{0, 1\}$

$$p(X) = \pi^X (1 - \pi)^{(1-X)}$$

Where π *is the probability of **success** (e.g. heads), and also the mean*

$$\mathbf{E}[X] = \pi \cdot 1 + (1 - \pi) \cdot 0 = \pi \qquad \mathbf{Var}[X] = \pi(1 - \pi)$$

Binomial Sum of N independent coinflips,

$$p(Y = k) = \binom{N}{k} \pi^k (1 - \pi)^{N-k}$$

With moments,

$$\mathbf{E}[Y] = N \cdot \pi \qquad \mathbf{Var}[Y] = N\pi(1 - \pi)$$

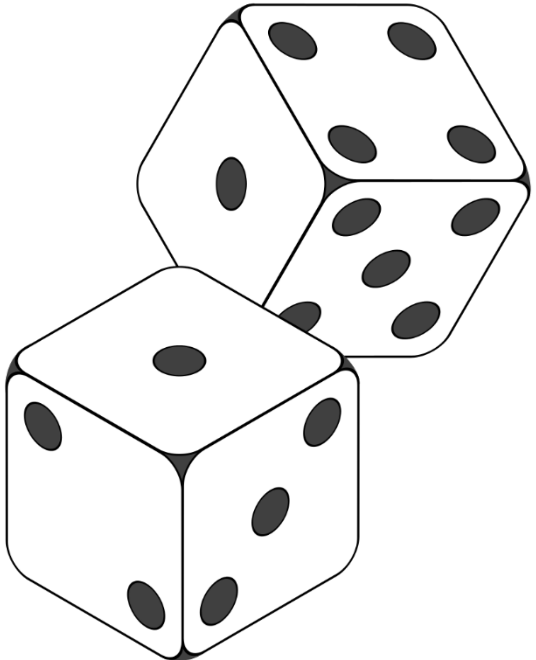
Moments of Useful Discrete Distributions

Count of *N* independent categorical RVs

$$p(x_1, \ldots, x_K) = \frac{N!}{x_1! x_2! \ldots x_K!} \prod_{k=1}^{K} \pi_k^{x_k}$$

Where RV X is a K-vector of counts and parameter $\pi \in [0,1]^K$ is a probability vector,

$$\sum_{k=1}^{K} \pi_k = 1$$



Marginal moments are given by,

$$\mathbf{E}[X_k] = N\pi_k \qquad \mathbf{Var}[X_k] = N\pi_k(1 - \pi_k)$$

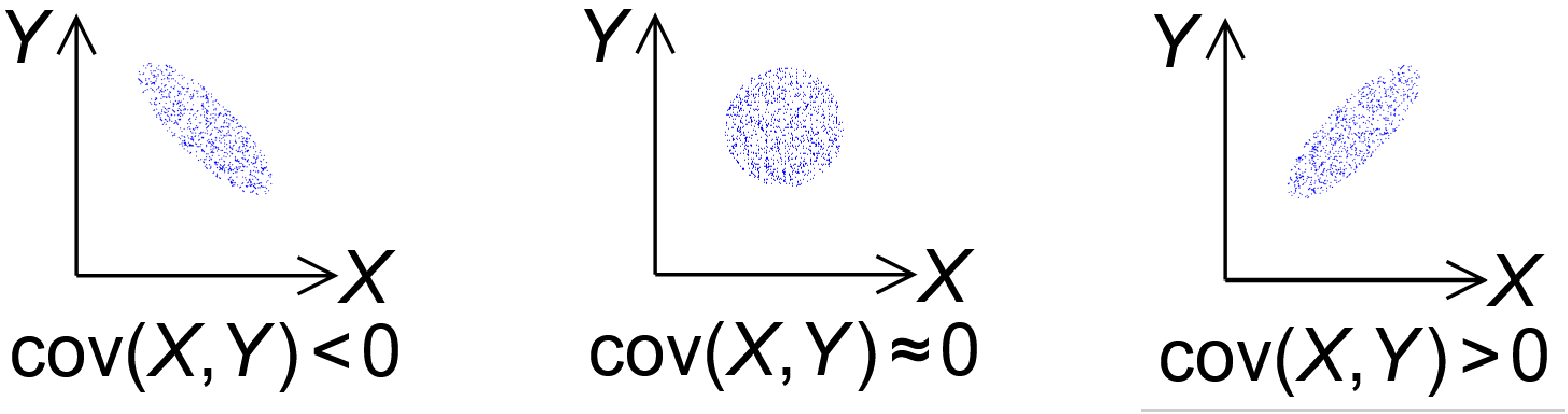
Moments are similar to Binomial, but over K outcomes

Covariance

Definition *The <u>covariance</u> of two RVs* X *and* Y *is defined as,*

$$\mathbf{Cov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$$

Measures the <u>linear relationship</u> between X and Y



Question *What is Cov(X,X)?*

Answer *Cov(X,X) = Var(X)*

Covariance

Lemma *For any two RVs* X *and* Y *,*

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X,Y)$$

e.g. variance is *<u>not a linear operator</u>.*

Proof $$\mathbf{Var}[X+Y] = \mathbf{E}[(X+Y-\mathbf{E}[X+Y])^2]$$

(Linearity of expectation) $$= \mathbf{E}[(X+Y-\mathbf{E}[X]-\mathbf{E}[Y])^2]$$

(Distributive property) $$= \mathbf{E}[(X-\mathbf{E}[X])^2 + (Y-\mathbf{E}[Y])^2 + 2(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$$

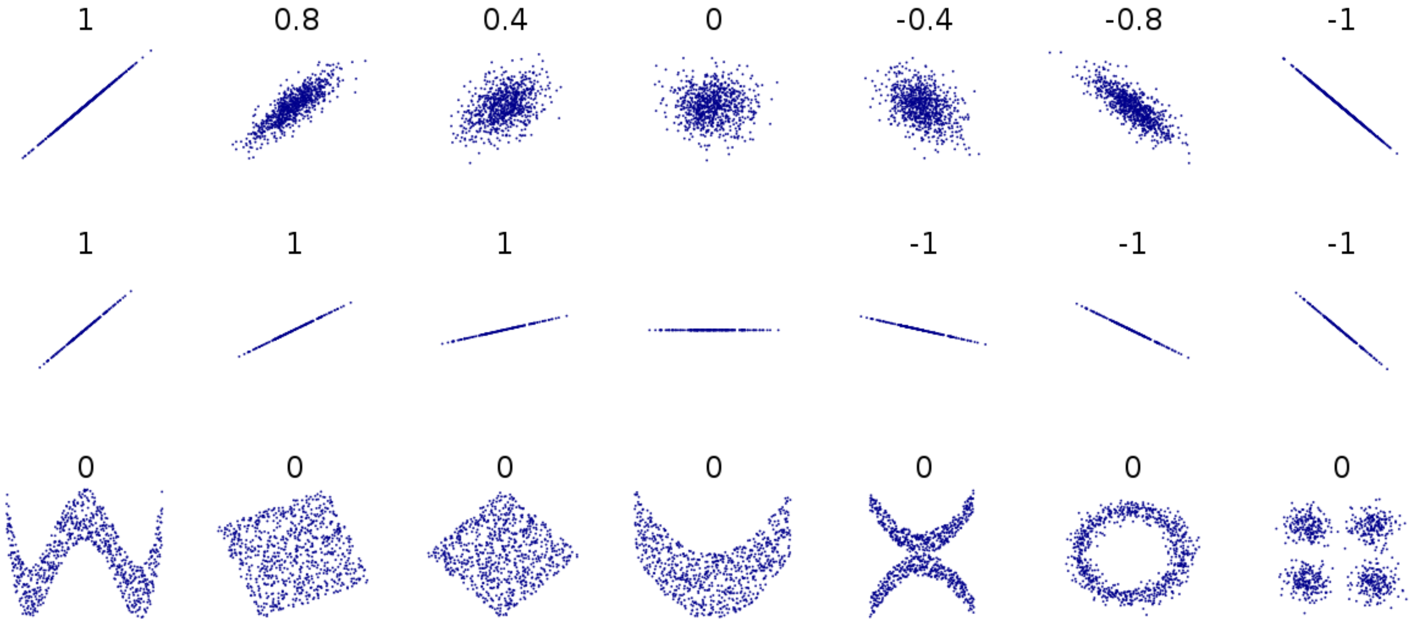
(Linearity of expectation) $$= \mathbf{E}[(X-\mathbf{E}[X])^2] + \mathbf{E}[(Y-\mathbf{E}[Y])^2] + 2\mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$$

(Definition of Var / Cov) $$= \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X,Y)$$

Correlation

Definition *The correlation of two RVs X and Y is given by,*

$$\mathbf{Corr}(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sigma_X \sigma_Y} \quad \textit{where} \quad \sigma_X = \sqrt{\mathbf{Var}(X)}$$



Like covariance, only expresses linear relationships!

Administrative Items

- HW1 Due Tonight @ 11:59pm
  - Do not have to answer Problem 2 (see Piazza)
- HW2 Out Tomorrow, Due Next Thursday 9/16 @ 11:59pm
  - 3 Questions, 2pts each
- Recall: Office Hours
  - Enfa : Monday @ 10:30, Gould-Simpson Rm 934, Desk #6 (hybrid)
  - Saiful : Tuesday @ 10:00, Gould-Simpson Rm 942 (hybrid)
  - Jason : Wednesday @ 10:00, (Zoom)

Intuition Check

<u>Question:</u> Roll two dice and let their outcomes be $X_1, X_2 \in \{1, \ldots, 6\}$ *for die 1 and die 2, respectively. Recall the definition of conditional probability,*

$$p(X_1 \mid X_2) = \frac{p(X_1, X_2)}{p(X_2)}$$

Which of the following are true?

a) $p(X_1 = 1 | X_2 = 1) > p(X_1 = 1)$

b) $p(X_1 = 1 | X_2 = 1) = p(X_1 = 1)$ Outcome of die 2 doesn't *affect* die 1

c) $p(X_1 = 1 | X_2 = 1) < p(X_1 = 1)$

Intuition Check

Question: *Let $X_1 \in \{1, \ldots, 6\}$ be outcome of die 1, as before. Now let $X_3 \in \{2, 3, \ldots, 12\}$ be the* sum of both dice. Which of the following are true?

a) $p(X_1 = 1 | X_3 = 3) > p(X_1 = 1)$

b) $p(X_1 = 1 | X_3 = 3) = p(X_1 = 1)$

c) $p(X_1 = 1 | X_3 = 3) < p(X_1 = 1)$

Only 2 ways to get $X_3 = 3$, each with equal probability:

$(X_1 = 1, X_2 = 2)$ *or* $(X_1 = 2, X_2 = 1)$

so

$$p(X_1 = 1 \mid X_3 = 3) = \frac{1}{2} > \frac{1}{6} = p(X_1 = 1)$$

- Expected Value
- Variance, Covariance, Correlation
- Dependence of Random Variables

Independence of RVs

Intuition…

Consider $P(B|A)$ where you want to bet on B

Should you pay to know A?

In general you would pay something for A if it changed your belief about B. In other words if,

$$P(B|A) \neq P(B)$$

Independence of RVs

Definition *Two random variables* X *and* Y *are* *<u>independent</u>* *if and only if,*

$$p(X = x \mid Y = y) = p(X = x)$$

for all values x and y, and we say $X \perp Y$. An equivalent definition is,

$$p(X = x, Y = y) = p(X = x)p(Y = y)$$

Definition RVs $X_1, X_2, \ldots, X_N$ are <u>mutually independent</u> if and only if,

$$p(X_1 = x_1, \ldots, X_N = x_N) = \prod_{i=1}^{N} p(X_i = x_i)$$

- Independence is *symmetric:* $X \perp Y \Leftrightarrow Y \perp X$

Independence of RVs

Definition *Two random variables* X *and* Y *are* <u>*conditionally independent*</u> *given* Z *if and only if,*

$$p(X = x \mid Y = y, Z = z) = p(X = x \mid Z = z)$$

for all values x, y, and z, and we say that $X \perp Y \mid Z$. Equivalently,

$$p(X = x, Y = y \mid Z = z) = p(X = x \mid Z = z)p(Y = y \mid Z = z)$$

- N RVs conditionally independent, given Z, if and only if:

$$p(X_1, \ldots, X_N \mid Z) = \prod_{i=1}^{N} p(X_i \mid Z)$$

Shorthand notation
Implies for all *x, y, z*

- Also symmetric: $X \perp Y \mid Z \Leftrightarrow Y \perp X \mid Z$

Independence and Moments

Theorem: *If* $X \perp Y$ *then* $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.

Proof:

$$\mathbf{E}[XY] = \sum_x \sum_y (x \cdot y) p(X = x, Y = y)$$

$$= \sum_x \sum_y (x \cdot y) p(X = x) p(Y = y)$$ **(Independence)**

$$= \left(\sum_x x \cdot p(X = x) \right) \left(\sum_y y \cdot p(Y = y) \right) = \mathbf{E}[X]\mathbf{E}[Y]$$ **(Linearity of Expectation)**

Example *Let* $X_1, X_2 \in \{1, \ldots, 6\}$ *be RVs representing the result of rolling two fair standard die.* *What is the mean of their product?*

$$\mathbf{E}[X_1 X_2] = \mathbf{E}[X_1]\mathbf{E}[X_2] = 3.5^2 = 12 \cdot \frac{1}{4}$$

Independence and Moments

Question: *What is the variance of their sum?*

$$\begin{aligned}
\mathbf{Var}[X_1 + X_2] &= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{Cov}(X_1, X_2) \\
&= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{E}[(X_1 - \mathbf{E}[X_1])(X_2 - \mathbf{E}[X_2])] \\
&= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\mathbf{E}[(X_1 - \mathbf{E}[X_1])]\mathbf{E}[(X_2 - \mathbf{E}[X_2])] \\
&= \mathbf{Var}[X_1] + \mathbf{Var}[X_2] + 2\left(\mathbf{E}[X_1] - \mathbf{E}[X_1]\right)\left(\mathbf{E}[X_2] - \mathbf{E}[X_2]\right) \\
&= \mathbf{Var}[X_1] + \mathbf{Var}[X_2]
\end{aligned}$$

*But wait… I thought variance was **not** a linear operator…*

Independence and Moments

Recall that for any two RVs X and Y variance is not a linear function,

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y] + 2\mathbf{Cov}(X,Y)$$

If X and Y are independent then they have zero covariance,

$$\mathbf{Cov}(X,Y) = 0$$

Thus variance is a linear operator for independent variables,

$$\mathbf{Var}[X+Y] = \mathbf{Var}[X] + \mathbf{Var}[Y]$$

And, for a collection of independent RVs $X_1, X_2, \ldots, X_N$ we have,

$$\mathbf{Var}(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} \mathbf{Var}(X_i)$$

Example: Independent Gaussian RVs

Let X and Y be independent Gaussian random variables with,

$$X \sim \mathcal{N}(\mu_x, \sigma_x^2) \qquad Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$$

What is the variance of their sum?

$$\mathbf{Var}(X+Y) = \mathbf{Var}(X) + \mathbf{Var}(Y) = \sigma_x^2 + \sigma_y^2$$

What is the mean of their product?

$$\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y] = \mu_x \mu_y$$

Suppose X and Y are **dependent**, what is the mean of their sum?

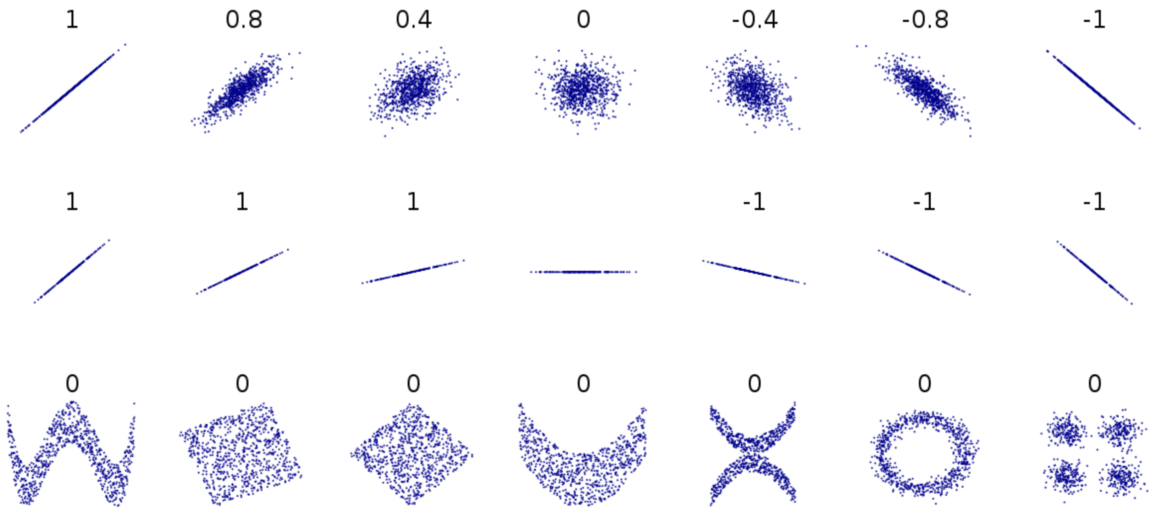
$$\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y] = \mu_x + \mu_y$$

Independence and Moments

From previous slide If X and Y are independent random variables, then:

$$\mathbf{Cov}(X, Y) = 0$$

The reverse is not true! $(\mathbf{Cov}(X, Y) = 0) \not\Rightarrow X \perp Y$



Example Let X be any RV and $Y=X^2$ then,

$$\mathbf{Cov}(X, Y) = 0$$

By direct calculation. Yet they are obviously dependant!

Moments of Continuous RVs

Replace all sums with integrals,

$$\mathbf{E}[X] = \int x p(x)\, dx \qquad \mathbf{Var}[X] = \int (x - \mathbf{E}[X])^2 p(x)\, dx$$

- All properties push through, as you would expect (e.g. law of total expectation, conditional expectation, etc.)
- In general you will not need to solve these integrals directly, you may use standard results for each distribution, e.g.

$$\mathbf{E}[X] = \int x \cdot \mathcal{N}(x \mid \mu, \sigma^2)\, dx = \mu$$

Review

We have covered a lot of ground on probability in short time…

Discrete Random Processes

- Definition of sample space / random events
- Axioms of probability
- Uniform probability of random event
- Random Variables
- Fundamental rules of probability (chain rule, conditional, law of total probability)

Probability Distributions

- Useful discrete probability mass functions’
- Introduction to continuous probability
- Useful probability density functions

Moments / Dependence

- Expected Value
- Linearity / Law of total expectation
- Variance, Covariance, Corr.
- Dependent / Independent RVs