Probability vs. Statistics

Probability

- Mathematical theory of uncertainty
- Parameters are usually known

Statistics

- The science of data analysis; techniques for making sense of data
- Parameters are usually unknown, and estimated from data

Random Variables

A baseball game is a random experiment

Sample statistics gathered:

- Number of runs, hits, and errors, per team
- Number of hits, walks, and outs, per player
- Number of strikeouts, walks, and pitches, per pitcher
- Etc.

These statistics are all examples of random variables.
What are other examples of random variables?

Intuition

- A random variable is a numerical outcome of an random experiment.
- If we flip a coin *n* times, there can be multiple random variables:
  - *H* can be the random variable indicating the number of heads
  - *T* = *n* - *H* can be the random variable indicating the number of tails
  - *X* can represent the longest sequence of consecutive heads
  - *Y* can represent the longest sequence of consecutive tails
- All these random variables are valued between 0 and *n*.

Definition

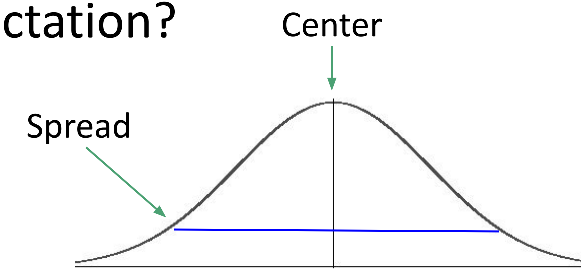
- A function is a mapping from a domain to a range.
  - E.g., $f(x,y) = x+y$ maps R^2 to R.
- A random variable is a function:
  - The domain of a random variable is (the set of subsets of) the universe of the experiment.
    - Ω = {HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}, if we flip a coin 3 times.
  - The range of a random variable is a subset of R, the reals.
- Example: Imagine flipping a fair coin 2 times.
  - Let *H* be the random variable representing the total number of heads.
  - Q: What is the domain? I.e. what are the possible outcomes? A: {HH, HT, TH, TT}.
  - Q: What is the range of *H*? I.e., what are the possible values of *H*? A: {0, 1, 2}.
  - Thus, *H* is a function from {HH, HT, TH, TT} to {0, 1, 2}.
  - $H(TT) = 0$; $H(HT) = H(TH) = 1$; $H(HH) = 2$

Probabilities

- Inversely, a random variable describes events: i.e., subsets of outcomes.
  - $H^{-1}(0)$ = {TT}; $H^{-1}(1)$ = {HT, TH}; $H^{-1}(2)$ = {HH}
- We associate a probability distribution with a random variable, based on the probabilities of the corresponding events.
- For example:
  - Pr[H = 0] = |{TT}| / |{HH, HT, TH, TT}| = ¼
  - Pr[H = 1] = |{HT, TH}| / |{HH, HT, TH, TT}| = ½
  - Pr[H = 2] = |{HH}| / |{HH, HT, TH, TT}| = ¼
- All probabilities are between 0 and 1, and together they sum to 1.

Random Variables Vary

- What makes random variables interesting is the fact that they vary!
- You can and will get different results running the same random experiment over and over again, as a result of the randomness.
- This randomness is described by a probability distribution, which is often summarized by its center and it spread.
- In other words, what value might you "expect"?
- And how might the actual value differ from this expectation?



Expectation

Expected Value

- Once we have a specified the range of a random variable, and its probability distribution, we can calculate the expected value of that random variable.
- Assume X is a random variable, x_i is an element of its range, and p_i is the probability of x_i. Then the expected value of X is calculated as follows:

$$E(X) = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + x_3 p_3 + \cdots + x_n p_n$$

- $E[H] = (0)\Pr[H = 0] + (1)\Pr[H = 1] + (2)\Pr[H = 2] = 0(¼) + 1(½) + (2)(¼) = 1$
    - If we flip a coin two times, we should expect to see one head.
- The expected value of a random variable is also called the mean (μ).

iClicker Question

If X is a random variable representing the result of one roll of a six-sided die, then what is the expected value of X?

A) 3
B) 3.5
C) 4

iClicker Answer

If X is a random variable for the number rolled on a six-sided die, then what is the expected value of X?

A) 3
B) 3.5
C) 4

$\frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5$

This is the expected value, even though we cannot roll a 3.5 ever!

Law of Large Numbers

- You can roll a die many times.
- Each time, you will see a 1, 2, 3, 4, 5, or 6.
- If you average the value you see across all your trials, as you run more and more experiments, this average will approach 3.5!

Trial	1	2	3	4	5	6	7	8	9	10	11	12
Value	1	4	3	3	5	6	2	3	4	1	6	4
Average	1	2.5	8/3	11/4	16/5	22/6	24/7	27/8	31/9	32/10	38/11	44/12

- In this experiment, the sample mean is 3.667.

Law of Large Numbers

- You can toss 2 coins many times.
- Each time, you will see 0, 1, or 2 heads.
- If you average the number of heads you see across all your trials, as you run more and more experiments, this average will approach 1!

Trial	1	2	3	4	5	6	7	8	9	10	11	12
Value	0	2	2	0	1	0	0	1	2	2	0	1
Average	0	1	4/3	1	1	5/6	5/7	6/8	10/9	12/10	12/11	13/12

- In this experiment, the sample mean is 1.083.

Variance

Recall Sample Variance

Assume *X* is a random variable; x_i is an element of its range; $\bar{x}$ is its sample mean. The sample variance of *X*, given a sample of size *N*, is calculated as follows:

$$s_X = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Actually, variance is not defined as the average of the squared deviations, but rather as the expected value of the squared deviations.

$$\sigma_X = \mathbb{E}_{x\sim P(X)}\left[(x - \mathbb{E}[X])^2\right]$$

Variance through the Prism of Expectation

- Once again, variance is the expected value of the squared deviations between the values of a random variable and its mean.
  - Variance[X] = $E[(X - E[X])^2]$
  - We can simplify this expression using algebra (and linearity of expectations).
  - Variance[X] = $E[X^2] - (E[X])^2$
- This second formula will yield the same result as the first, and it is usually easier to work with!
- The variance of the number of heads in two fair coin flips is:
  - $E[X^2] = (¼)0^2 + (½)1^2 + (¼)2^2 = 1.5$
  - $E[X]^2 = 1^2 = 1$
  - $E[X^2] - E[X]^2 = 1.5 - 1 = ½$

iClicker Question

If X is a random variable representing the value, then what is the variance of X?

- $E[X^2] = (½)0^2 + (½)1^2 = ½$
- $(E[X])^2 = (½)^2 = ¼$
- $E[X^2] - (E[X])^2 = ½ - ¼ = ¼$

iClicker Question

If X is a random variable representing the result of one roll of a six-sided die, then what is the variance of X?

- $E[X^2] = (⅙)1^2 + (⅙)2^2 + (⅙)3^2 + (⅙)4^2 + (⅙)5^2 + (⅙)6^2 = 15.167$
- $E[X]^2 = 3.5^2 = 12.25$
- $E[X^2] - E[X]^2 = 15.167 - 12.25 \sim 2.9167$

Recall Sample Variance

Assume *X* is a random variable; x_i is an element of its range; $\bar{x}$ is its sample mean. The sample variance of *X*, given a sample of size *N*, is calculated as follows:

$$s_X = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

Trial	1	2	3	4	5	6	7	8	9	10	11	12
Value	1	4	3	3	5	6	2	3	4	1	6	4
Squared Differences	6.25	0.25	0.25	0.25	2.25	6.25	2.25	0.25	0.25	6.25	6.25	0.25
Avg of Sq Diff	6.25	3.25	2.25	1.75	1.85	2.583	2.536	2.25	2.028	2.45	2.795	2.583

- In this experiment, the r.v. is the value of the roll of a six-sided die.
- The expected value of this r.v. is 3.5, and it's variance is 2.9167.

Recall Sample Variance

Assume *X* is a random variable; x_i is an element of its range; $\bar{x}$ is its sample mean. The sample variance of *X*, given a sample of size *N*, is calculated as follows:

$$s_X = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Trial	1	2	3	4	5	6	7	8	9	10	11	12
Value	0	2	2	0	1	0	0	1	2	2	1	1
Squared Differences	1	1	1	1	0	1	1	0	1	1	0	0
Avg of Sq Diff	1	1	1	1	.8	.833	.85714	.75	.777	.8	.7272	667.

- In this experiment, the r.v. is the number of heads in two coin flips.
- The expected value of this r.v. is 1, and it's variance is 0.5.

ICA

Expectation and Variance

Let X be a random variable representing the result of the sum of the rolls of two six-sided dice.

- What is the expected value of X?
- What is the variance of X?

Expectation

Let X be a random variable representing the result of the sum of the rolls of two six-sided dice.

$$E[X] = (1/36)2 + (1/18)3 + (1/12)4 + (1/9)5 + (5/36)6 + (1/6)7 + (5/36)8 + (1/9)9 + (1/12)10 + (1/18)11 + (1/36)12 = 7$$

Variance

Let *X* be a random variable representing the result of the sum of the rolls of two six-sided dice.

$E[X^2] = (1/36)2^2 + (1/18)3^2 + (1/12)4^2 + (1/9)5^2 + (5/36)6^2 + (1/6)7^2 + (5/36)8^2 + (1/9)9^2 + (1/12)10^2 + (1/18)11^2 + (1/36)12^2 \sim 54.83$

$(E[X])^2 = (7)^2 = 49$

$\text{Variance}[X] = E[X^2] - (E[X])^2 \sim 54.83 - 49 = 5.83$

A Technical Aside

The Linearity of Expectation

- We introduced the concept of expectation in the last lecture.
- Let X_1 be a random variable representing the first die roll.
  - $E[X_1] = 3.5$
- Let X_2, X_3, X_4, and X_5 represent additional rolls of the dice.
  - $E[X_1 + X_2 + X_3 + X_4 + X_5] = E[5X_1] = 5E[X_1] = 17.5$
- More generally:
  - $E[aY_1 + bY_2] = E[aY_1] + E[bY_2] = aE[Y_1] + bE[Y_2]$
- This rule is called the linearity of expectation.

The Nonlinearity of Variance

- Remember that variance is the square of standard deviation.
  - You can pull constants out of the expectation formula.
  - You can pull constants out of variance as well,
    but you must square them when you do so.
  - $\mathrm{Var}[aX + b] = a^2\mathrm{Var}[X]$
- Additional properties:
  - $\mathrm{Var}[X - Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$
  - $\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$, if X and Y are independent:
    I.e., Variance is linear when X and Y are independent

Two Random Variables

A baseball game is a random experiment

Sample statistics gathered:

- Number of pitches thrown, number of hits
- Number of balls thrown, number of base runners
- Length of game, number of runs
- Etc.

These statistics are more examples of random variables.

What are other examples of random variables?

Recall: Independent Events

- Recall the following, from the definition of conditional probability:
  - P(A and B) = P(A|B)P(B)
  - P(A and B) = P(B|A)P(A)
- Mathematically, two events are independent if:
  - P(A|B) = P(A)
  - P(B|A) = P(B)
- Consequently, two events are independent if P(A and B) = P(A)P(B).

Independent Random Variables

- Recall the following, from the definition of conditional probability:
  - P(X and Y) = P(X|Y)P(Y)
  - P(X and Y) = P(Y|X)P(A)
- Mathematically, two random variables are independent if:
  - P(X|Y) = P(X)
  - P(Y|X) = P(Y)
- Consequently, two random variables are independent if P(X and Y) = P(X)P(Y).

Rolling Two Three-Sided Dice

JOINT DISTRIBUTION

X \ Y	1	2	3	MARGINAL DIST'N
1	1/9 (1,1)	1/9 (1,2)	1/9 (1,3)	1/3
2	1/9 (2,1)	1/9 (2,2)	1/9 (2,3)	1/3
3	1/9 (3,1)	1/9 (3,2)	1/9 (3,3)	1/3
MARGINAL DIST'N	1/3	1/3	1/3	

MARGINAL DIST'N

Prob[X = 1] =
Prob[X = 1, Y = 1] +
Prob[X = 1, Y = 2] +
Prob[X = 1, Y = 3] =
1/9 + 1/9 + 1/9 = 1/3

Law of Total Probability

Rolling Two Three-Sided Dice

JOINT DISTRIBUTION

	1	2	3
1	1/9	1/9	1/9
2	1/9	1/9	1/9
3	1/9	1/9	1/9

SUMS

	1	2	3
1	2	3	4
2	3	4	5
3	4	5	6

iClicker Question

If *X* is a random variable representing the result of the sum of the rolls of two three-sided dice, then what is the expected value of *X*?

$$E(x) = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + x_3 p_3 + \cdots + x_n p_n$$

$E[X] = (1/9)\ 2 + (2/9)\ 3 + (3/9)\ 4 + (2/9)\ 5 + (1/9)\ 6 = 4$

Rolling Two Loaded Three-Sided Dice

JOINT DISTRIBUTION

	1	2	3	MARGINAL DIST’N
1	⅙	1/12	1/12	1/3
2	1/12	⅙	1/12	1/3
3	1/12	1/12	⅙	1/3
MARGINAL DIST’N	1/3	1/3	1/3	

Rolling Two Loaded Three-Sided Dice

JOINT DISTRIBUTION

	1	2	3
1	⅙	1/12	1/12
2	1/12	⅙	1/12
3	1/12	1/12	⅙

SUMS

	1	2	3
1	2	3	4
2	3	4	5
3	4	5	6

iClicker Question

If X is a random variable representing the result of the sum of the rolls of two three-sided dice, then what is the expected value of X?

$$E(x) = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + x_3 p_3 + \cdots + x_n p_n$$

E[X] = (1/6) 2 + (1/6) 3 + (1/3) 4 + (1/6) 5 + (1/6) 6 = 4

Rolling Two Loaded Three-Sided Dice

JOINT DISTRIBUTION

	1	2	3	MARGINAL DIST’N
1	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{3}$
2	$\frac{1}{6}$	0	0	$\frac{1}{6}$
3	0	0	$\frac{1}{2}$	$\frac{1}{2}$
MARGINAL DIST’N	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{2}$	

Rolling Two Loaded Three-Sided Dice

JOINT DISTRIBUTION

	1	2	3
1	⅙	⅙	0
2	⅙	0	0
3	0	0	½

SUMS

	1	2	3
1	2	3	4
2	3	4	5
3	4	5	6

iClicker Question

If *X* is a random variable representing the result of the sum of the rolls of two three-sided dice, then what is the expected value of *X*?

$$E(x) = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + x_3 p_3 + \cdots + x_n p_n$$

E[*X*] = (⅙) 2 + (⅓) 3 + (½) 6 = 4.333

Joint Distribution

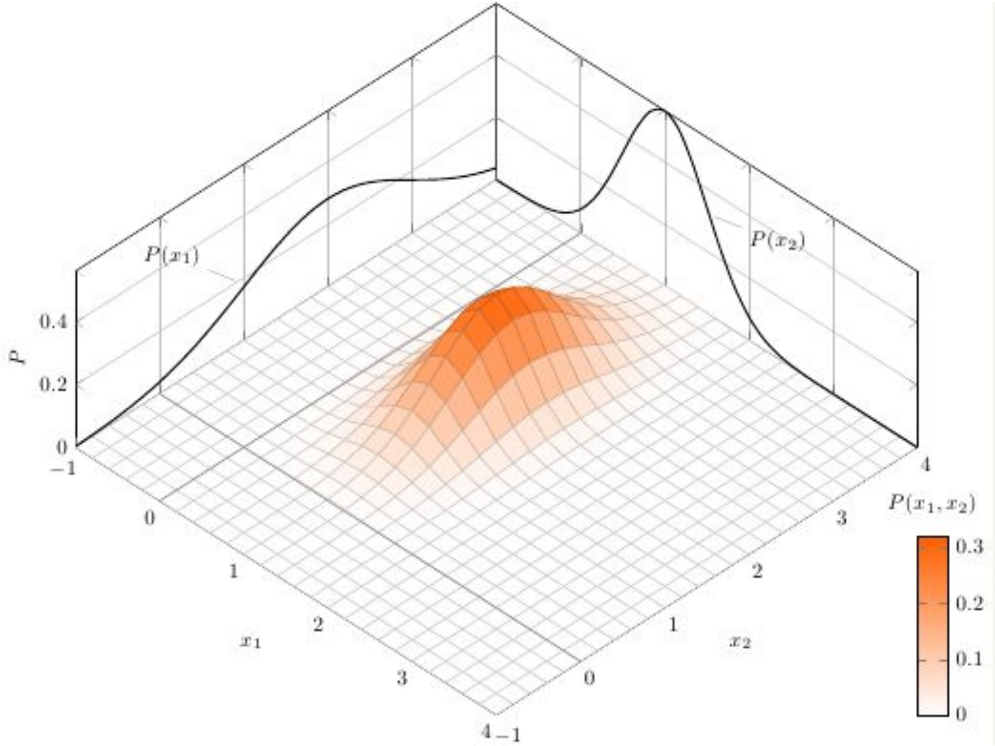


Image source

Joint Distribution

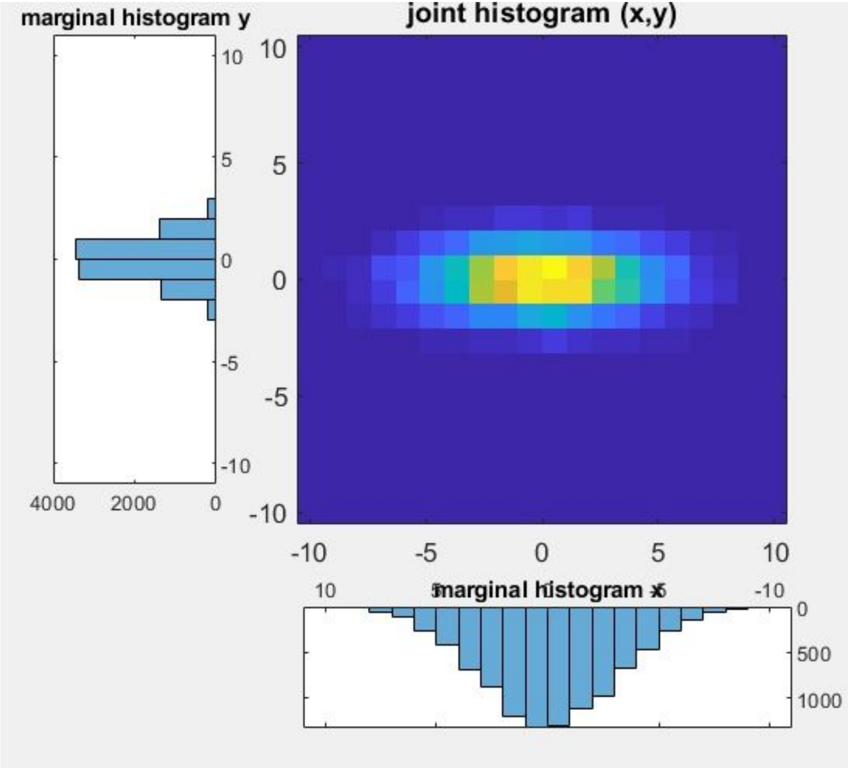


Image source

Covariance

Recall Sample Covariance

Assume *X* is a random variable, x_i is an element of its range, and μ_x is its mean.
Assume *Y* is a random variable, y_i is an element of its range, and μ_y is its mean.
The sample covariance of *X* and *Y*, given a sample of size *N*, is calculated as follows:

$$s_{XY} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

Actually, covariance is not defined as the average product of the deviations, but rather as the expected product:

$$\sigma_{XY} = \mathbb{E}_{(x,y)\sim P(X,Y)}\left[(x - \mathbb{E}[X])(y - \mathbb{E}[Y])\right]$$

Covariance through the Prism of Expectation

- Once again, covariance is the product of the expected values of the deviations between the values of two random variables and their means.
    - Covariance[X, Y] = E[(X - E[X])(Y - E[Y])]
    - We can simplify this through some algebra.
    - Covariance[X, Y] = E[XY] - E[X]E[Y]
- This second formula will yield the same result as the first, and it is usually easier to work with!
- Corr[X, Y] = Cov[X, Y] / $\sigma_X \sigma_Y$

Rolling Two Loaded Three-Sided Dice

JOINT DISTRIBUTION

	1	2	3	MARGINAL DIST'N
1	⅙	0	0	⅙
2	0	⅓	0	⅓
3	0	0	½	½
MARGINAL DIST'N	⅙	⅓	½	

iClicker Question

What is the covariance of these two loaded three-sided dice?

- $E[XY] = \frac{1}{6}(1) + \frac{1}{3}(4) + \frac{1}{2}(9) = 6$
- $E[X] = E[Y] = \frac{1}{6}(1) + \frac{1}{3}(2) + \frac{1}{2}(3) = 2\frac{1}{3}$
- $\mathrm{Cov}[X, Y] = E[XY] - E[X]E[Y] = 6 - (2\frac{1}{3})^2 \approx 0.555$

In this example, $E[XY] = E[X^2]$ and $E[X]E[Y] = (E[X])^2$,
so $\mathrm{Cov}[X, Y] = E[XY] - E[X]E[Y] = E[X^2] - (E[X])^2 = \mathrm{Var}[X]$.
Likewise, $\mathrm{Cov}[X, Y] = \mathrm{Var}[Y]$.

iClicker Question

What is the correlation of these two loaded three-sided dice?

- $E[X^2] = E[Y^2] = \frac{1}{6}(1)^2 + \frac{1}{3}(2)^2 + \frac{1}{2}(3)^2 = 6$
- $E[X] = E[Y] = \frac{1}{6}(1) + \frac{1}{3}(2) + \frac{1}{2}(3) = 2\frac{1}{3}$
- $Var[X] = Var[Y] = 6 - (2\frac{1}{3})^2 \approx 0.555$

$Corr[X, Y] = Cov[X, Y] / \sigma_X\sigma_Y{}^2 \approx 0.555 / (\sqrt{0.555}\sqrt{0.555}) = 1$

More generally, whenever $Cov[X, Y] = Var[X] = Var[Y]$,
$Corr[X, Y] = Cov[X, Y] / \sigma_X\sigma_Y = Var[X] / \sigma_X\sigma_X = Var[X] / Var[X] = 1$.

Rolling Two Loaded Three-Sided Dice

JOINT DISTRIBUTION

	2	4	6	MARGINAL DIST'N
1	⅙	0	0	⅙
2	0	⅓	0	⅓
3	0	0	½	½
MARGINAL DIST'N	⅙	⅓	½	

iClicker Question

What is the covariance of these two loaded three-sided dice?

- $E[XY] = \frac{1}{6}(2) + \frac{1}{3}(8) + \frac{1}{2}(18) = 12$
- $E[X] = \frac{1}{6}(1) + \frac{1}{3}(2) + \frac{1}{2}(3) = 2\frac{1}{3}$
- $E[Y] = \frac{1}{6}(2) + \frac{1}{3}(4) + \frac{1}{2}(6) = 4\frac{2}{3}$
- $Cov[X, Y] = E[XY] - E[X]E[Y] = 12 - (2\frac{1}{3})(4\frac{2}{3}) \approx 1.111$

iClicker Question

What is the correlation of these two loaded three-sided dice?

- $E[X^2] = \frac{1}{6}(1)^2 + \frac{1}{3}(2)^2 + \frac{1}{2}(3)^2 = 6$
- $E[X] = \frac{1}{6}(1) + \frac{1}{3}(2) + \frac{1}{2}(3) = 2\frac{1}{3}$
- $Var[X] = 6 - (2\frac{1}{3})^2 \approx 0.555$
- $E[Y^2] = \frac{1}{6}(2)^2 + \frac{1}{3}(4)^2 + \frac{1}{2}(6)^2 = 24$
- $E[Y] = \frac{1}{6}(2) + \frac{1}{3}(4) + \frac{1}{2}(6) = 4\frac{2}{3}$
- $Var[Y] = 24 - (4\frac{2}{3})^2 \approx 2.222$

$Corr[X, Y] = Cov[X, Y] / \sigma_X\sigma_Y \approx 1.111 / (\sqrt{0.555}\ \sqrt{2.222}) = 1$

JOINT DISTRIBUTION	Rain tomorrow (.75)	No rain tomorrow (.25)
Rain today (.6)	.55	.05
No rain today (.4)	.2	.2

- The covariance of rain today and rain tomorrow is:
  - $E[XY] = .55(1)(1) + .05(1)(0) + .2(0)(1) + .2(0)(0) = .55$
  - $E[X] = P(X = 1) = .6$
  - $E[Y] = P(Y = 1) = .75$
  - $E[XY] - E[X]E[Y] = .55 - .45 = .1$
- The variance of rain today and rain tomorrow is:
  - $E[X^2] = P(X = 1)(1^2) = .6$
  - $E[Y^2] = P(Y = 1)(1^2) = .75$
  - $Var[X] = E[X^2] - (E[X])^2 = .6 - (.6)^2 = .24$
  - $Var[Y] = E[Y^2] - (E[Y])^2 = .75 - (.75)^2 = .1875$
- $Corr[X, Y] = Cov[X, Y] / \sigma_X\sigma_Y = .1 / \sqrt{(.24)}\sqrt{(.1875)} = .4714$

Flipping Two Loaded Coins

JOINT DISTRIBUTION

	0	1	MARGINAL DIST’N
0	⅔	0	⅔
1	0	⅓	⅓
MARGINAL DIST’N	⅔	⅓	

iClicker Question

What is the covariance of these two loaded three-sided dice?

- $E[XY] = \frac{2}{3}(0) + \frac{1}{3}(1) = \frac{1}{3}$
- $E[X] = E[Y] = \frac{2}{3}(0) + \frac{1}{3}(1) = \frac{1}{3}$
- $\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = \frac{1}{3} - (\frac{1}{3})^2 \approx 0.222$

In this example, $E[XY] = E[X^2]$ and $E[X]E[Y] = (E[X])^2$,
so $\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = E[X^2] - (E[X])^2 = \text{Var}[X]$.
Likewise, $\text{Cov}[X, Y] = \text{Var}[Y]$.

iClicker Question

What is the correlation of these two loaded three-sided dice?

- $E[X^2] = E[Y^2] = ⅔\,(0)^2 + ⅓\,(1)^2 = ⅓$
- $E[X] = E[Y] = ⅔\,(0) + ⅓\,(1) = ⅓$
- $Var[X] = Var[Y] = ⅓ - (⅓)^2 \approx 0.222$

$Corr[X, Y] = Cov[X, Y] / \sigma_X \sigma_Y \approx 0.222 / (\sqrt{0.222}\,\sqrt{0.222}) = 1$

More generally, whenever $Var[X] = Var[Y] = Cov[X, Y]$,
$Corr[X, Y] = Cov[X, Y] / \sigma_X \sigma_Y = Var[X] / \sigma_X \sigma_X = Var[X] / Var[X] = 1$.

Flipping Two Loaded Coins

JOINT DISTRIBUTION

	0	1	MARGINAL DIST'N
0	0	⅔	⅔
1	⅓	0	⅓
MARGINAL DIST'N	⅓	⅔	

iClicker Question

What is the covariance of these two loaded three-sided dice?

- $E[XY] = \frac{1}{3}(0) + \frac{2}{3}(0) = 0$
- $E[X] = \frac{2}{3}(0) + \frac{1}{3}(1) = \frac{1}{3}$
- $E[Y] = \frac{1}{3}(0) + \frac{2}{3}(1) = \frac{2}{3}$
- $Cov[XY] = E[XY] - E[X]E[Y] = 0 - (\frac{1}{3})(\frac{2}{3}) \approx -0.222$

iClicker Question

What is the correlation of these two loaded three-sided dice?

- $E[X^2] = E[Y^2] = \frac{2}{3}(0)^2 + \frac{1}{3}(1)^2 = \frac{1}{3}$
- $\text{Var}[X] = \frac{1}{3} - (\frac{1}{3})^2 \approx 0.222$
- $\text{Var}[Y] = \frac{2}{3} - (\frac{2}{3})^2 \approx 0.222$

$\text{Corr}[X, Y] = \text{Cov}[X, Y] / \sigma_X\sigma_Y \approx -0.222 / (\sqrt{0.222}\ \sqrt{0.222}) = -1$

More generally, whenever $\text{Var}[X] = \text{Var}[Y] = -\text{Cov}[X, Y]$,
$\text{Corr}[X, Y] = \text{Cov}[X, Y] / \sigma_X\sigma_Y = -\text{Var}[X] / \sigma_X\sigma_X = -\text{Var}[X] / \text{Var}[X] = -1.$

Extras

Rolling Two Loaded Three-Sided Dice

JOINT DISTRIBUTION

	1	2	3	MARGINAL DIST'N
1	⅙	0	0	⅙
2	0	0	½	½
3	0	⅓	0	⅓
MARGINAL DIST'N	⅙	½	⅓	

iClicker Question

What is the covariance of these two loaded three-sided dice?

- $E[XY] = (\frac{1}{6})\ 1 + (\frac{1}{3} + \frac{1}{2})\ 6 = 5\frac{1}{6}$
- $E[X] = (\frac{1}{6})\ 1 + (\frac{1}{2})\ 2 + (\frac{1}{3})\ 3 = 2\frac{1}{6}$
- $E[Y] = (\frac{1}{6})\ 1 + (\frac{1}{3})\ 2 + (\frac{1}{2})\ 3 = 2\frac{1}{3}$
- $\mathrm{Cov}[XY] = E[XY] - E[X]E[Y] = 5\frac{1}{6} - (2\frac{1}{6})(2\frac{1}{3}) \approx 0.111$

iClicker Question

What is the covariance of these two loaded three-sided dice?

- $E[X^2] = (\frac{1}{6})\, 1^2 + (\frac{1}{2})\, 2^2 + (\frac{1}{3})\, 3^2 = 5\frac{1}{6}$
- $E[X] = (\frac{1}{6})\, 1 + (\frac{1}{2})\, 2 + (\frac{1}{3})\, 3 = 2\frac{1}{6}$
- $\mathrm{Var}[X] = 5\frac{1}{6} - (2\frac{1}{6})^2 \approx 0.47222$
- $\mathrm{SD}[X] = \sqrt{\mathrm{Var}[X]} \approx 0.687$

- $E[Y^2] = (\frac{1}{6})\, 1^2 + (\frac{1}{3})\, 2^2 + (\frac{1}{2})\, 3^2 = 6$
- $E[Y] = (\frac{1}{6})\, 1 + (\frac{1}{3})\, 2 + (\frac{1}{2})\, 3 = 2\frac{1}{3}$
- $\mathrm{Var}[Y] = 6 - (2\frac{1}{3})^2 \approx 0.555$
- $\mathrm{SD}[Y] = \sqrt{\mathrm{Var}[Y]} \approx 0.745$

$\mathrm{Corr}[X,Y] = \mathrm{Cov}[X,Y] / \sigma\, \sigma \approx 0.111 / (0.687)(0.745) \approx 0.217$