

The Importance of Summarizing Distributions

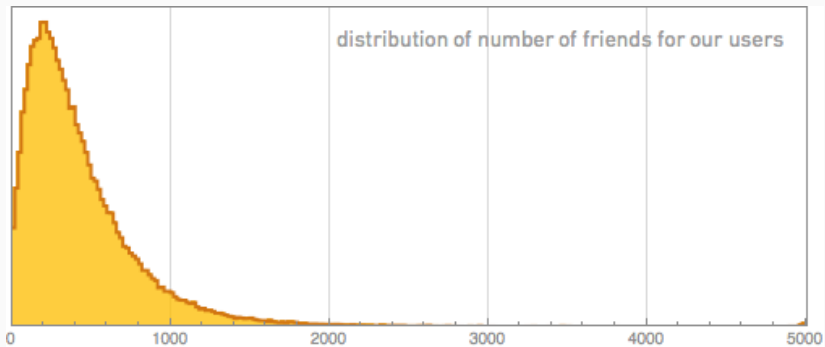
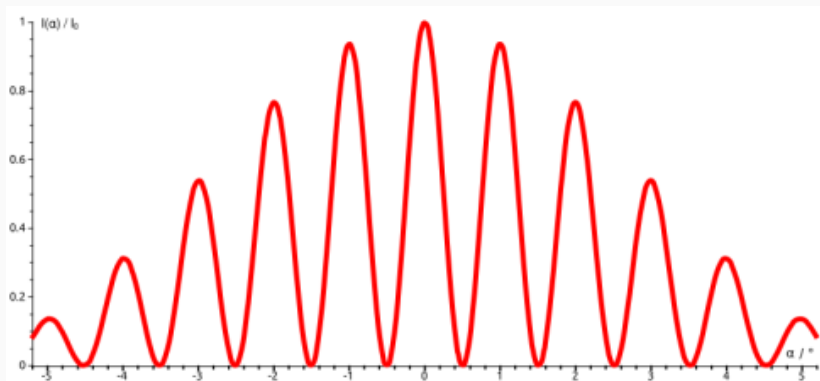
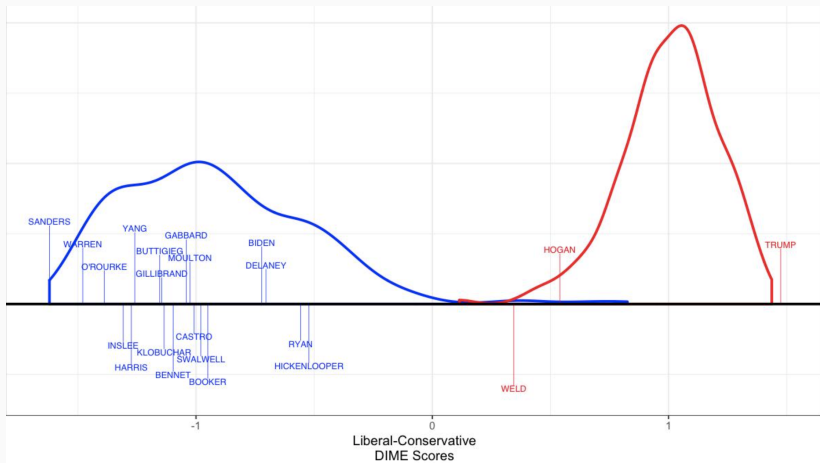


Image: stephenwolfram.com. Data Science of the Facebook World.





Source: Database on Ideology, Money in Politics, and Elections (DIME).

THE DOWNSIDE OF FLEXIBILITY

Highly flexible objects present obstacles.

- Cannot specify density for every value
- Cannot communicate or reason about every value
- Cannot estimate density at every value given data

Why not restrict to distributions in a parametric family?

- The real world isn't that neat!
- "Agnostic" approach: minimal assumptions on the model

WHY SUMMARIZE?

Summary tools are the hooks we use to estimate, reason about, and communicate about random variables.

FOUNDATIONAL SUMMARY TOOLS

One Random Variable:

- Expectation: represents the center of a distribution
- Variance: represents the dispersion of a distribution

Two Random Variables

- Covariance: a generalization of variance
- Correlation: an intuitive measure of linear association

Review of Operators and Functions

REVIEW OF OPERATORS AND FUNCTIONS

Function $f(x) = x^2$



$f(X)$ is a random variable

Operator E



$E[X]$ is a real number

Reading: Summarizing Distributions

READING ASSIGNMENT

Read *Foundations of Agnostic Statistics*, page 44 to the top of page 47, stopping after you read theorem 2.1.5.

Lightboard: Expectation of a Discrete Random Variable

EXPECTATION OF A DISCRETE RANDOM VARIABLE

EXPECTATION FOR DISCRETE RANDOM VARIABLES

$$E[X] = \sum_{x \in \text{supp}[X]} xf(x)$$

Ex 1: $X \sim \text{Bernoulli}(\alpha)$

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

$$E[X] = 0 \cdot (1 - \alpha) + 1 \cdot \alpha = \alpha$$

Ex 2: $Y \sim \text{Geometric}(\beta)$

Let J_i be the event printer jams day i .

$P(J_i) = \beta$, all ind.

Let Y be days until first jam

$$P(Y = 1) = P(J_1) = \beta$$

$$P(Y = 2) = P(J_1^C, J_2) = (1 - \beta)\beta$$

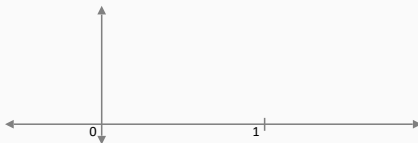
$$P(Y = 3) = (1 - \beta)^2\beta$$

$$f(x) = \begin{cases} (1 - \beta)^{x-1}\beta, & x \in \{1, 2, \dots\} \\ 0, & \text{otherwise} \end{cases}$$

$$E[Y] = \sum_{x=1}^{\infty} x(1 - \beta)^{x-1}\beta = 1/\beta$$

Lightboard: Expectation of a Continuous Random Variable

EXPECTATION OF A CONTINUOUS RANDOM VARIABLE



EXPECTATION FOR CONTINUOUS RANDOM VARIABLES

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

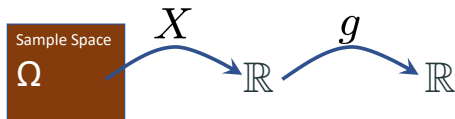
Ex 1: $X \sim \text{Uniform}(a, b)$

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_{-\infty}^a x \cdot 0dx + \int_a^b x \frac{1}{b-a}dx + \int_b^{\infty} x \cdot 0dx \\ &= 0 + \frac{x^2}{2} \frac{1}{b-a} \Big|_a^b + 0 \\ &= \frac{b^2-a^2}{2} \frac{1}{b-a} = \frac{(b-a)(b+a)}{2} \frac{1}{b-a} \\ &= \frac{a+b}{2} \end{aligned}$$

Law of the Unthinking Statistician

LAW OF THE UNTHINKING STATISTICIAN (LOTUS)



LAW OF THE UNTHINKING STATISTICIAN (LOTUS)



Proof sketch:

X is a DRV with pmf f
 $g(X)$ is a DRV with pmf,
 $f'(y) = P(g(X) = y) =$
 $\sum_{x \in \mathbb{R}: g(x)=y} f(x)$
want: $E[g(X)]$
 $E[g(X)] = \sum_x g(x)f(x)$

$$\begin{aligned} E[g(X)] &= \sum_y y f'(y) \\ &= \sum_y \sum_{x \in \mathbb{R}: g(x)=y} y f(x) \\ &= \sum_y \sum_{x \in \mathbb{R}: g(x)=y} g(x) f(x) \\ &= \sum_{x \in \text{Supp}[X]} g(x) f(x) \end{aligned}$$

Reading: Linearity of Expectation

READING ASSIGNMENT

Read *Foundations of Agnostic Statistics*, pages 47–50, stopping at 2.1.2.

Lightboard: Linearity of Expectation

LINEARITY OF EXPECTATION

Assume RV C with pmf

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

then C is a constant.

$$E[C] = \sum_x x f(x) = c f(c) = c$$

*usually write c for RV.

Let X, Y be RVs. $a \in \mathbb{R}$

$$E[aX] = \sum_{x \in \text{Supp}[X]} a x f_X(x) =$$

$$a \sum_{x \in \text{Supp}[X]} x f_X(x) = a E[X]$$

$$E[X + Y] =$$

$$\sum_{x \in X(\Omega), y \in Y(\Omega)} (x + y) f(x, y) =$$

$$\sum_x \sum_y x f(x, y) + \sum_y \sum_x y f(x, y)$$

$$= \sum_x x \sum_y f(x, y) +$$

$$\sum_y y \sum_x f(x, y)$$

$$= \sum_x x f_X(x) + \sum_y y f_Y(y)$$

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

Reading: Moments and Variance

READING ASSIGNMENT

Read *Foundations of Agnostic Statistics*, section 2.1.2.

Lightboard: Moments and Variance

MOMENTS AND VARIANCE

(raw) moments:

1. $E[X]$ 2. $E[X^2]$ 3. $E[X^3]$...

Define variance.

$$\begin{aligned}V[aX] &= E[(aX)^2] - E[aX]^2 = \\&= a^2 E[X^2] - (aE[X])^2 = \\&= a^2 E[X^2] - a^2 E[X]^2 = a^2 V[X] \\V[X + c] &= \\E[(X + c - E[X + c])^2] &= \\E[(X + c - E[X] - E[c])^2] &= \\E[(X - E[X])^2] &= V[X]\end{aligned}$$

Reading: Mean Squared Error (MSE)

READING ASSIGNMENT

Read *Foundations of Agnostic Statistics*, section 2.1.3.

Lightboard: Mean Squared Error (MSE) - Lightboard

THE EXPECTED VALUE AND MSE

Proof that $E[X]$ minimizes MSE

Covariance and Correlation

Measuring Linear Dependency

UNDERSTANDING RELATIONSHIPS

Most of the really important questions in data science are about *relationships*.

- Bitterness of coffee

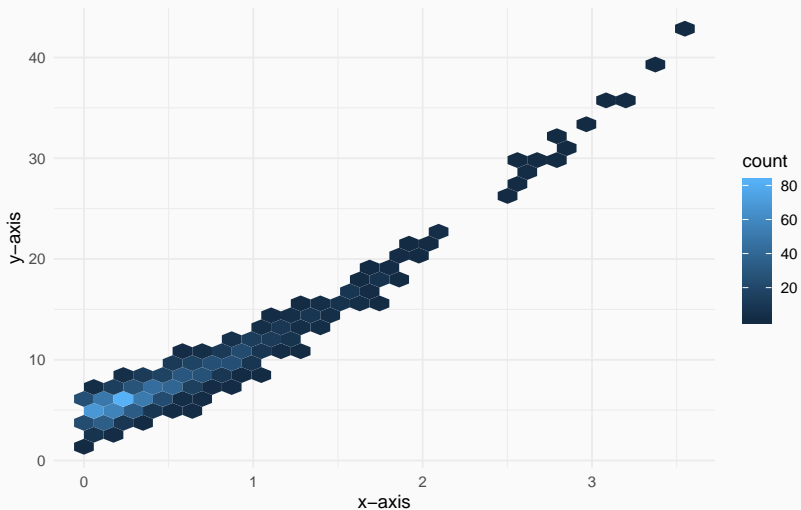
UNDERSTANDING RELATIONSHIPS

Most of the really important questions in data science are about *relationships*.

- Bitterness of coffee
- Roasting temperature and time

**How can we make sense of the wide
variety of relationships among
variables?**

THE JOINT DISTRIBUTION



MEASURING DEPENDENCY

A good first question about two random variables:

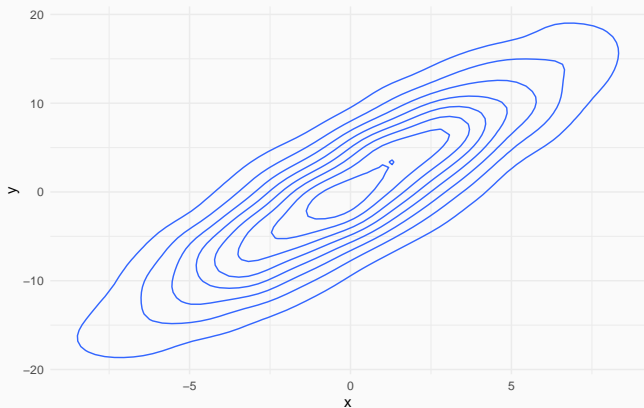
How strong is the relationship?

Two common tools to answer this question:

- Covariance
- Correlation

INTUITION FOR COVARIANCE

$$\text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$



Reading Assignment

READING ASSIGNMENT

Note: This is a reading call, we're just placing it here for organization. Read *Foundations of Agnostic Statistics*, pages 59–62, stopping at Correlation.

Variance of Sums

VARIANCE OF SUMS

Insert content from old section of course: 4.10 Variance of Sums

Alternative Formula for Covariance

ALTERNATIVE FORMULA FOR COVARIANCE

$$\text{cov}[X, Y] = E[(X - E[X])(Y - E[Y])]$$

Learnosity

LEARNOSITY: $V[X - Y]$

You have just read theorem 2.2.3, which says, in part:

Variance of sums

$$V[X + Y] = V[X] + 2\text{cov}[X, Y] + V[Y]$$

1. If X and Y are independent — that is, $\text{cov}[X, Y] = 0$ — which is larger? (a) $V[X - Y]$; (b) $V[X + Y]$; (c) They are the same. (d) We don't have enough information.
2. If X and Y are not independent — that is, $\text{cov}[X, Y] \neq 0$ — which is larger? (a) $V[X - Y]$; (b) $V[X + Y]$; (c) They are the same. (d) We don't have enough information.

LEARNOSITY: $V[X - Y]$ (CONT.)

1. Under what circumstances would it be possible for $V[X]$ and $V[Y]$ to “cancel out”?
2. Notice that $V[X + (-1 * X)] = V[X - X] = V[0] = 0$.
Use this to find a simple formula for $cov[X, X]$.

Properties of Covariance Lightboard

PROPERTIES OF COVARIANCE

For random variables X, Y, Z, W and $a, b \in \mathbb{R}$,

LIGHTBOARD: LINEARITY OF COVARIANCE

Note: This is Lightboard. We're just placing it here for organization. For random variables X, Y, Z, W and constants a, b ,

$$\begin{aligned}\text{cov}[aX, bY] &= E[aXbY] - E[aX]E[bY] = \\ &abE[XY] - abE[X]E[Y] = ab\text{cov}[X, Y]\end{aligned}$$

$$\begin{aligned}\text{cov}[X + Y, Z] &= E[(X + Y) \cdot Z] - E[X + Y]E[Z] \\ &= E[XZ + YZ] - (E[X] + E[Y])E[Z] = \\ &E[XZ] + E[YZ] - E[X]E[Z] - E[Y]E[Z] \\ &= \text{cov}[X, Z] + \text{cov}[Y, Z]\end{aligned}$$

$$\text{b.s.a } \text{cov}[X, Y + Z] = \text{cov}[X, Y] + \text{cov}[X, Z]$$

$$\begin{aligned}\text{cov}[X + Y, Z + W] &= \text{cov}[X, Z + W] + \text{cov}[Y, Z + W] \\ &= \text{cov}[X, Z] + \text{cov}[X, W] + \text{cov}[Y, Z] + \text{cov}[Y, W]\end{aligned}$$

Correlation

CORRELATION IS RESCALED COVARIANCE

Definition 2.2.5

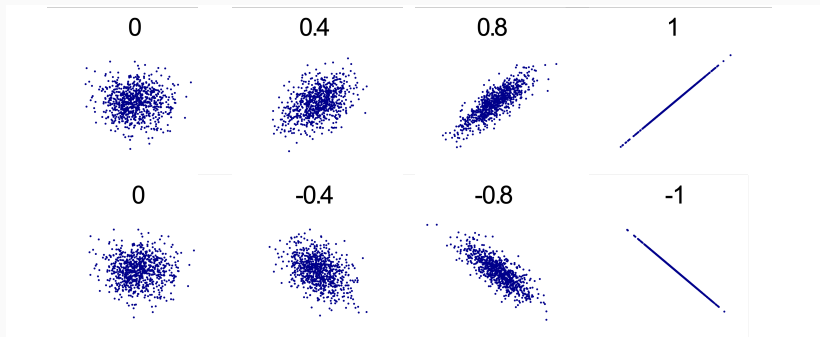
Correlation is a rescaled derivative of covariance that captures the *linear dependence* between two random variables.

$$\rho[X, Y] = \frac{\text{cov}[X, Y]}{\sigma[X]\sigma[Y]}$$

Society has a plain-language usage of *correlation* that is probably closer in usage to *covariance* than to *correlation*.

CORRELATION MEASURES LINEAR DEPENDENCY

Correlation for example distributions:

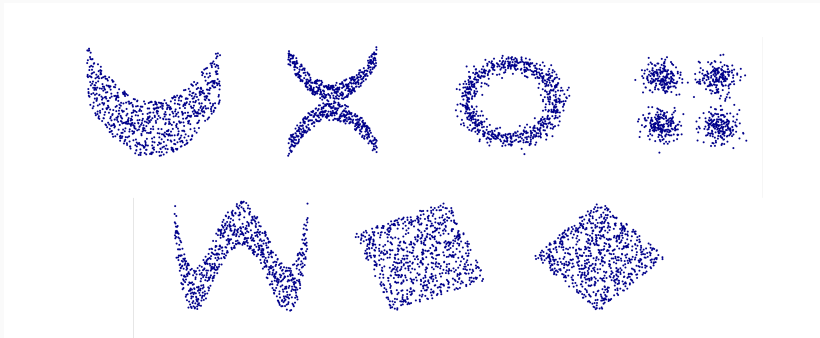


Source:

en.wikipedia.org/wiki/Correlation_and_dependence

NOT ALL DEPENDENCY IS LINEAR

Example distributions with zero correlation:



Source:

en.wikipedia.org/wiki/Correlation_and_dependence

Reading: Correlation

READING ASSIGNMENT

Read pages 62-65, stopping before you get to example 2.2.9.

Theorem 2.2.7

The third and fourth bullet points can be more clearly stated.

- If a and b are either both positive or both negative,

$$\rho[aX + c, bY + d] = \rho[X, Y]$$

- If a and b have opposite signs,

$$\rho[aX + c, bY + d] = -\rho[X, Y]$$

Covariance, Correlation, and Independence

INDEPENDENT VARIABLES HAVE ZERO CORRELATION

Independence $\implies \rho = 0$

If X and Y are independent random variables, then:

- $E[XY] = E[X]E[Y]$
- $\text{cov}[X, Y] = 0$
- $\rho[X, Y] = 0$
- $V[X + Y] = V[X - Y] = V[X] + V[Y]$

INDEPENDENT VARIABLES HAVE ZERO CORRELATION

Wrap-Up of Correlation

WRAP-UP OF CORRELATION

Note: This is a solo-lecture. We are just placing this here for organization.

Note: This will be a README. We're just placing it here for organization.