Week 3

Summarizing Distributions

Paul Laskowski and Alex Hughes

January 12, 2023

UC Berkeley, School of Information

The Importance of Summarizing Distributions

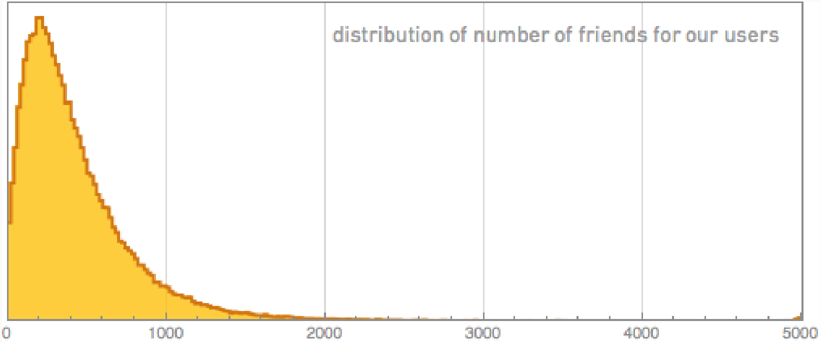
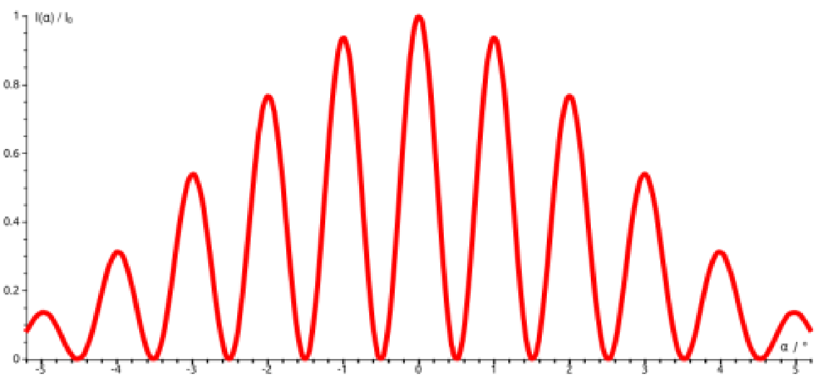
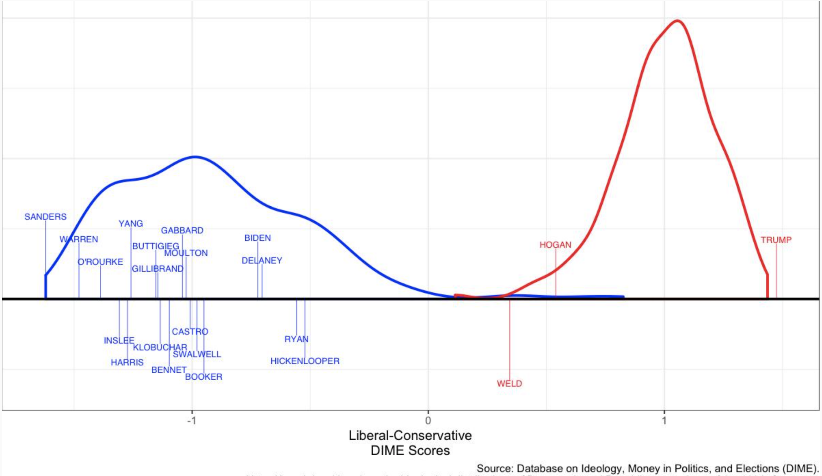


Image: stephenwolfram.com. Data Science of the Facebook World.



I(α) / I₀
1
0.8
0.6
0.4
0.2
0
-5
-4
-3
-2
-1
0
1
2
3
4
5
α / °



SANDERS
WARREN
O'ROURKE
YANG
BUTTIGIEG
GILLIBRAND
GABBARD
MOULTON
BIDEN
DELANEY
HOGAN
TRUMP
INSLEE
HARRIS
KLOBUCHAR
BENNET
CASTRO
SWALWELL
BOOKER
RYAN
HICKENLOOPER
WELD
-1
0
1
Liberal-Conservative
DIME Scores
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

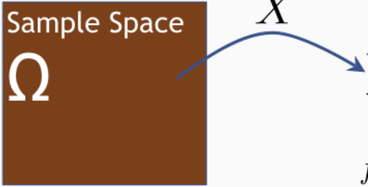
Two Random Variables

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

Review of Operators and Functions

Review of Operators and Functions

Function $f(x) = x^2$



X $\qquad$ $\mathbb{R}$ $\xrightarrow{f}$ $\mathbb{R}$

$f(X)$ is a random variable

Operator E

$X \xrightarrow{\mathrm{E}} \mathbb{R}$

$\mathrm{E}[X]$ 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

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

Ex 1 : $X \sim$ Bernoulli(α)

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

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

Ex 2: $Y \sim$ Geometric(β)

Let J_i be the event printer jams day i.

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

Let Y be days until first jam

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

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

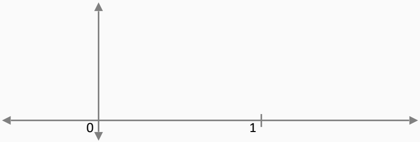
$P(Y = 3) = (1 - \beta)^2\beta$

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

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

Lightboard: Expectation of a Continuous Random Variable

Expectation of a Continuous Random Variable



Expectation for Continuous Random Variables

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

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

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

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

$= \int_{-\infty}^{a} x \cdot 0dx + \int_{a}^{b} x\frac{1}{b-a}dx + \int_{b}^{\infty} x \cdot 0dx$

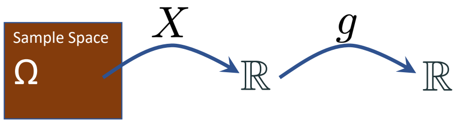
$= 0 + \frac{x^2}{2}\frac{1}{b-a}|_a^b + 0$

$= \frac{b^2-a^2}{2}\frac{1}{b-a} = \frac{(b-a)(b+a)}{2}\frac{1}{b-a}$

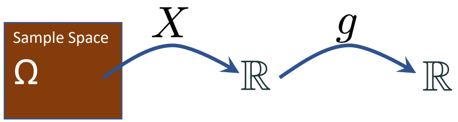
$= \frac{a+b}{2}$

Law of the Unthinking Statistician

Law of the Unthinking Statistician (LOTUS)



Law of the Unthinking Statistician (LOTUS)



Proof sketch:

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

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: $E[g(X)]$

$E[g(X)] = \sum g(x) f(x)$

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

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

then C is a constant.

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

*usually write c for RV.

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

$E[aX] = \sum_{x \in Supp[X]} axf_X(x) = a \sum_{x \in Supp[X]} xf_X(x) = aE[X]$

$E[X + Y] = \sum_{x \in X(\Omega), y \in Y(\Omega)} (x + y) f(x, y) = \sum_x \sum_y xf(x, y) + \sum_y \sum_x yf(x, y)$

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

$= \sum_x xf_X(x) + \sum_y yf_Y(y)$

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

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.

$$V[aX] = E[(aX)^2] - E[aX]^2 =$$
$$a^2E[X^2] - (aE[X])^2 =$$
$$a^2E[X^2] - a^2E[X]^2 = a^2V[X]$$
$$V[X + c] =$$
$$E\Big[(X + c - E[X + c])^2\Big] =$$
$$E\Big[(X + c - E[X] - E[c])^2\Big] =$$
$$E\Big[(X - E[X])^2\Big] = V[X]$$

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 $\mathrm{E}[X]$ 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

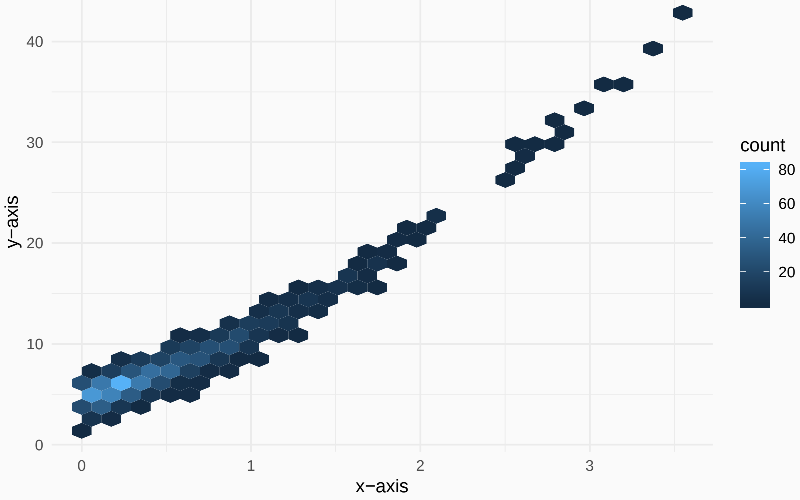
Understanding Relationships

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:

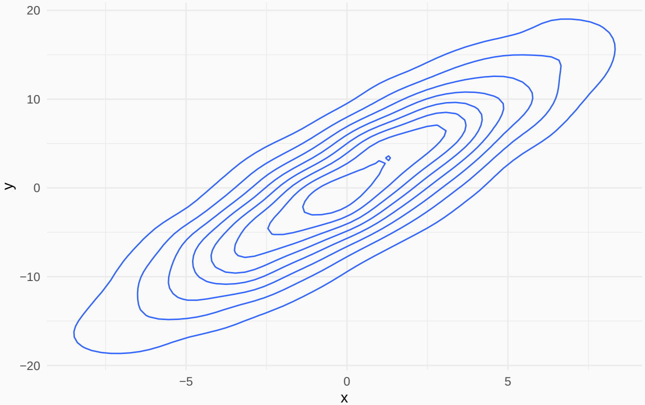
How strong is the relationship?

Two common tools to answer this question:

- Covariance
- Correlation

Intuition for Covariance

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



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

$$\operatorname{cov}[X, Y] = \mathrm{E}\Big[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])\Big]$$

Learnosity

Learnosity: $\mathsf{V}[X - Y]$

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

Variance of sums

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

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

Learnosity: $\mathbf{V}[X - Y]$ (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,

$cov[aX, bY] = E[aXbY] - E[aXE[BY] =$
$abE[XY] - abE[X]E[Y] = abcov[X, Y]$

$cov[X + Y, Z] = E[(X + Y) \cdot Z] - E[X + Y]E[Z]$
$= E[XZ + YZ] - (E[X] + E[Y])E[Z] =$
$E[XZ] + E[YZ] - E[X]E[Z] - E[Y]E[Z]$
$= cov[X, Z] + cov[Y, Z]$

b.s.a $cov[X, Y + Z] = cov[X, Y] + cov[X, Z]$

$cov[X + Y, Z + W] = cov[X, Z + W] + cov[Y, Z + W]$
$= cov[X, Z] + cov[X, W] + cov[Y, Z] + cov[Y, W]$

Correlation

Correlation is Rescaled Covariance

Definition 2.2.5

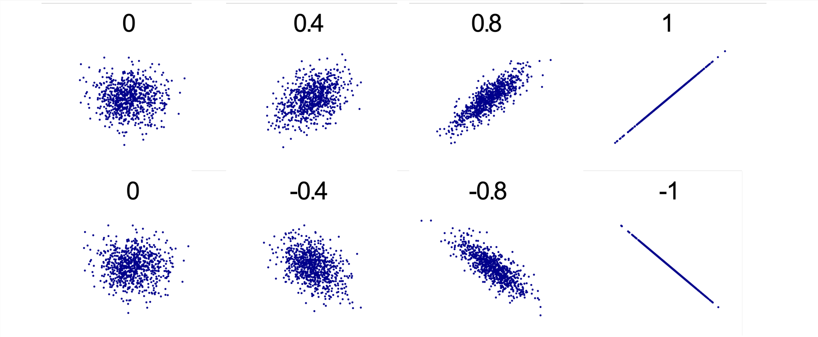
Correlation is a rescaled derivative of covariance that captures the *linear dependence* between two random variables.

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

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

Correlation Measures Linear Dependency

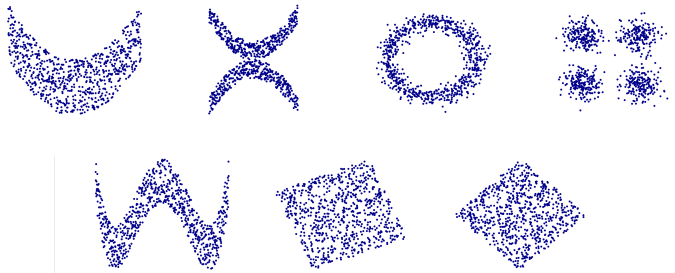
Correlation for example distributions:



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:

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

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.

Coding Activity

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