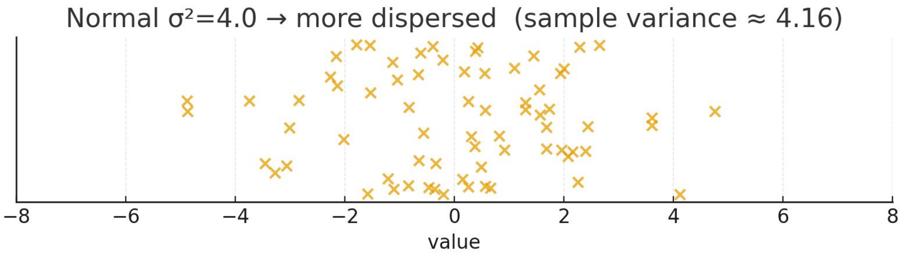
Causal Inference Recitation-Week 2

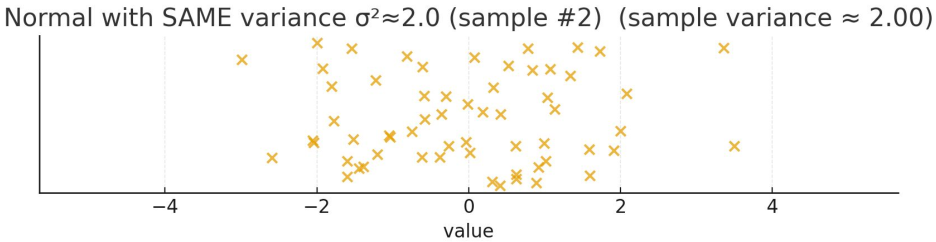
Alex Zhang

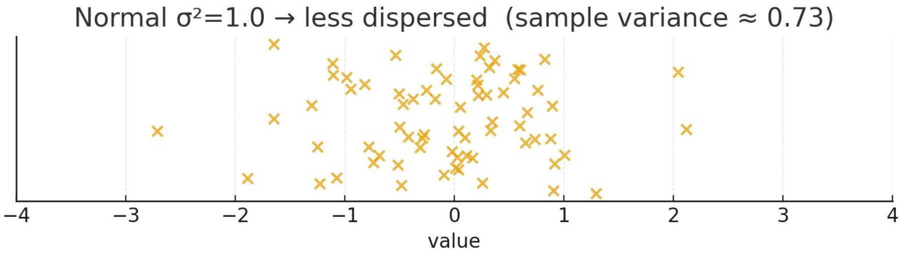
Sep 12, 2025

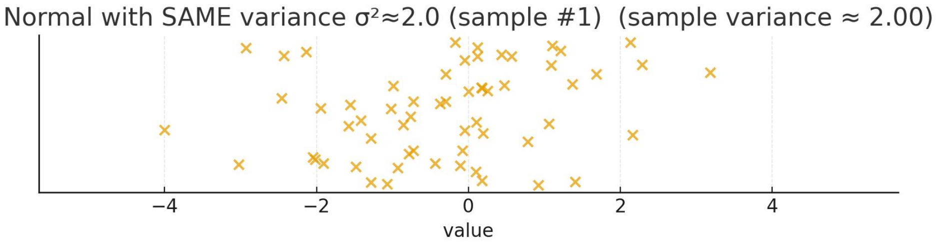
Parameter 2: Variance:

The variance of a R.V. is a parameter <u>measuring how disperse the distribution is</u>.









Parameter 2: Variance:

The variance of a R.V. is a parameter measuring how disperse the distribution is.

$$\sigma_X^2 = Var(X) = E[(X - E[X])^2]$$

The units of the variance are the units of X squared.

Alternative Variance Formula:

$$Var(X) = E[X^2] - E[X]^2$$

$$\begin{aligned} E[(X - E[X])^2] &= E[X^2 - 2XE[X] + E[X]^2] && \text{(algebra)} \\ &= E[X^2] - E[2XE[X]] + E[X]^2 && \text{(property ii)} \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 && \text{(property i)} \\ &= E[X^2] - E[X]^2 && \blacksquare \end{aligned}$$

T/F: It is possible for variance to have a negative value

Joint Distribution:

Def: A joint distribution characterizes how two or more variables are simultaneously distribution.

Forms:

- **Joint PMF** for discrete variables: $P(X = x, Y = y)$.
- **Joint PDF** for continuous variables: $f(x, y)$

Example: If we have a pair of dice, we can represent the exact roll as a random vector (X,Y) where X is the value of one dice and Y the value of the other。

Then, for any $x \in \{1, 2, \ldots, 6\}$ and $y \in \{1, 2, \ldots, 6\}$,

$$P(X = x, Y = y) = \frac{1}{36}$$

Joint Distribution to Marginal Distribution:

Def: Deriving the distribution of one individual rv from a joint distribution

Calculation (Discrete): The probability of one rv taking on a given value is the sum of the probabilities of all random vectors in which that variable takes on that value.

For 2 variables, if X takes on k values: $P(Y=y)=\sum_{i=1}^{k} P(Y=y, X=x_i)$

In our pair of dice example, what's the chance of the Y die equalling 4?

$$\begin{aligned} P(Y=4) &= \sum_{i=1}^{6} P(Y=4, X=i) \\ &= \frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36} \\ &= \frac{1}{6} \end{aligned}$$

Covariance:

Def: Covariance is an important parameter of joint distribution, and it captures whether 2 variables move together or not.

-It has the units: (units of X) x (units of Y).

Calculation (Discrete): The probability of one rv taking on a given value is the sum of the probabilities of all random vectors in which that variable takes on that value.

$$
\begin{aligned}
Cov(X,Y) &= E[(X - E[X])(Y - E[Y])] \\
&= E[XY] - E[X]E[Y]
\end{aligned}
$$

$$
\begin{aligned}
Cov(X,Y) &= E[(X - E[X])(Y - E[Y])] \\
&= E[XY - XE[Y] - YE[X] + E[X]E[Y]] \\
&= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] \\
&= E[XY] - 2E[X]E[Y] + E[X]E[Y] \\
&= E[XY] - E[X]E[Y] \quad \blacksquare
\end{aligned}
$$

Covariance:

T/F: For two real-valued random variables X and Y with finite variances, it is possible that $\mathrm{Cov}(X, Y) = 0$.

Example: Covariance $= 0$ with independent coin flips

Let $A, B \in \{0, 1\}$ encode fair coin flips (1=heads, 0=tails).
Independence $\Rightarrow P(A = a, B = b) = \frac{1}{4}$ for all $a, b \in \{0, 1\}$.

Expectations

$$E[A] = 1 \cdot \tfrac{1}{2} + 0 \cdot \tfrac{1}{2} = \tfrac{1}{2}, \qquad E[B] = \tfrac{1}{2}.$$

Compute $E[AB]$ (discrete expectation)

$$E[AB] = \sum_{a,b \in \{0,1\}} ab\, P(A = a, B = b) = (1 \cdot 1)\tfrac{1}{4} + (1 \cdot 0)\tfrac{1}{4} + (0 \cdot 1)\tfrac{1}{4} + (0 \cdot 0)\tfrac{1}{4} = \tfrac{1}{4}.$$

(Equivalently: $AB = 1$ **iff** $A = B = 1$; otherwise $AB = 0$.
So $E[AB] = P(A = 1, B = 1) = \frac{1}{4}$.)

Covariance

$$\mathrm{Cov}(A, B) = E[AB] - E[A]E[B] = \tfrac{1}{4} - \left(\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right) = 0.$$

Covariance:

Properties of Covariance

Let X, Y, and Z be rv's and a and b scalars. Then

- $Cov(X, Y) = Cov(Y, X)$
- $Cov(X, a) = 0$
- $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$
- $Cov(a + bX, Y) = bCov(X, Y)$
- $Cov(X, X) = Var(X)$
- $Var(a + bX) = b^2 Var(X)$

Correlation

Correlation between X and Y is defined as:

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

a normalized version of covariance.

Correlation is unitless, and for any X and Y, $|Corr(X, Y)| \in [0, 1]$.

- If $Corr(X, Y) = 0$, we say the 2 variables are uncorrelated. And if 2 variables has a covariance of 0, they will be uncorrelated.
- $Corr(X, Y) = 1$ iff $Y = a + bX$ for some scalar a and some positive scalar b. (Since a negative b means $Corr(X, Y) = -1$.

Correlation and Covariance

Correlation (and by extension) does not imply causation. Even if two variables are observed moving together in the real world, one does not necessarily cause the other.

In similar way, causation also doesn't imply correlation.

CRIME

When Ice Cream Sales Rise, So Do Homicides. Coincidence, or Will Your Next Cone Murder You?

BY JUSTIN PETERS

JULY 09, 2013 • 2:59 PM

Conditional Distributions

The **conditional distribution** tells us the likelihood of an outcome(s) for one variable if we know the outcome of another variable.

- For discrete random variables X and Y, and x_i such that $P(X = x_i) > 0$

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}$$

- For continuous X, Y, we'll have the conditional PDF:

$$f(y|x) = \frac{f(x, y)}{f(x)}$$

Conditional Expectations

With conditional distributions, we can consider **conditional expectations**, which are our "best guess" at the value of one R.V. given what we know about another R.V..

- For discrete R.V.s, where Y takes on k values:

$$E[Y|X=x] = \sum_{j=1}^{k} y_j P(Y=y_j|X=x_i)$$

$$\begin{aligned} E[Z|Y=6] &= \sum_{z=1}^{12} zP(Z=z|Y=6) \\ &= 1 * P(Z=1|Y=6) + 2 * P(Z=2|Y=6) + \\ &\quad ... + 12 * P(Z=12|Y=6) \\ &= 0 + (7+8+9+10+11+12)\frac{1}{6} \\ &= 9.5 \end{aligned}$$

Z = the sum of 2 die rolls, X and Y. We know that Y = 6.

- For continuous R.V., we have the definition:

$$E[Y|X=x] = \int_{r=1}^{\infty} yf(y|x)dy$$

It is important to note that:

- Expectations conditional on unrealized R.V., $E[Y|X]$, are functions and are random.
- A conditional expectation evaluated at a specific value of the conditioning variable, $E[Y|X=x]$ is a parameter (a specific number).

Properties of Conditional Expectation:

- Let X, Y, and Z be rv's. For any functions g and h,
  (i) $E[g(X) + h(X)Y|X] = g(X) + h(X)E[Y|X]$
  (ii) $E[Y + Z|X] = E[Y|X] + E[Z|X]$

Law of Iterated Expectations

We can go from conditional expectations to unconditional expectations using an extremely important tool - **the Law of Iterated Expectations**

$$E[Y] = E[E[Y|X]]$$

Quick example (two fair dice):

Roll two **independent** fair dice. Let X be the first die (values $1-6$), Y the second, and $Z = X + Y$.

For each $y \in \{1, \ldots, 6\}$,

$$\mathbb{E}[Z \mid Y = y] = \mathbb{E}[X + y \mid Y = y] = \mathbb{E}[X] + y = 3.5 + y.$$

Averaging over the six equally likely y's,

$$\mathbb{E}[Z] = \frac{4.5 + 5.5 + 6.5 + 7.5 + 8.5 + 9.5}{6} = 7,$$

which also matches $\mathbb{E}[X] + \mathbb{E}[Y] = 3.5 + 3.5 = 7$.

NYU

Conditional Variance

The **conditional variance** of Y given X is a function of X:

$$Var(Y|X) = E[(Y - E[Y|X])^2|X] \\ = E[Y^2|X] - E[Y|X]^2$$

$$\begin{aligned} Var(Y|X) &= E[(Y - E[Y|X])^2|X] \\ &= E[Y^2 - 2YE[Y|X] + E[Y|X]^2|X] \\ &= E[Y^2|X] - 2E[Y|X]E[Y|X] + E[Y|X]^2 \\ &= E[Y^2|X] - 2E[Y|X]^2 + E[Y|X]^2 \\ \Rightarrow Var(Y|X) &= E[Y^2|X] - E[Y|X]^2 \end{aligned}$$

Properties of Conditional Variance

Let X and Y be R.V., then for any functions g and h:

$$Var[g(X) + h(X)Y|X] = h^2(X)Var[Y|X]$$

Independence

R.V. X and Y are independent, deonted $X \perp Y$, if, for any sets A and B:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

$$\begin{aligned} P(X = 6, Y = 6) &= P(X = 6) * P(Y = 6) \\ &= \frac{1}{6} * \frac{1}{6} \\ &= \frac{1}{36} \end{aligned}$$

For 2 distinct dices, X and Y

We can say that their rolls are unrelated (or the dice are independent) if it can do this with any group of outcomes for each die.