Gov 2000 - 4. Multiple Random Variables

Matthew Blackwell

September 29, 2015

Where are we? Where are we going?

- We described a formal way to talk about uncertain outcomes, probability.
- We've talked about how to use that framework to characterize and summarize the uncertainty in one random variable.
- What about relationships between variables? How do we represent these?
- Need to talk about multiple r.v.s at the same time.
- **Remember!** We're learning about the features of some underlying distribution of the data, which don't get to observe. In the coming weeks, we'll talk about how to estimate these features using data.

Why multiple random variables?

- We already looked at the distribution of each variable separately.
- Now, how do we talk about two variables together?
- $\rightsquigarrow$ how do we summarize the relationship between these variables?

1/ Distributions of Multiple Random Variables

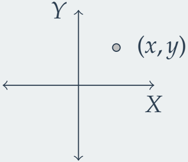
Joint and Conditional Probabilities

- Joint probability of two events, A and B: $\mathbb{P}(A \cap B)$
- Conditional probability of A given B:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

- We want to merge these concepts with the concept of random variables.
  - How does the distribution of one variable change as we change the value of another variable?
- Examples:
  - Does changing the negativity of advertising change the distribution of turnout?
  - Does changing economic conditions change the distribution of support for barriers to trade?
  - Does changing electoral system change the distribution of party ideologies?

Joint distributions



- The joint distribution of two r.v.s, X and Y, describes what pairs of observations, (x, y) are more likely than others.
- According to the DGP, should random samples from this joint distribution be:
  - clustered in a cloud?
  - roughly oriented along a line?
  - some other way?

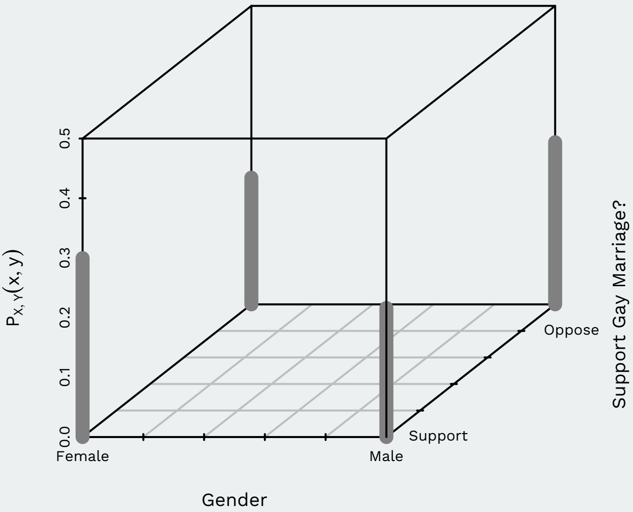
Discrete r.v.s

- X and Y both be discrete random variables.
- What is the probability that $X = x$ and $Y = y$ both occur?
- **Definition**: The can be fully described by the **joint probability mass function** of (X, Y) is:

$$f_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}(\{X = x\} \cap \{Y = y\})$$

- Properties of a joint p.m.f.:
    - $f_{X,Y}(x, y) \geq 0$ (probabilities can't be negative)
    - $\sum_x \sum_y f_{X,Y}(x, y) = 1$ (something must happen)
    - $\sum_x$ is short-hand for "sum over all possible values of X"

Example: Gay marriage and gender



Example: Gay marriage and gender

	Favor Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$
Female $X = 1$	0.3	0.21
Male $X = 0$	0.22	0.27

- Joint p.m.f. can be summarized in a cross-tab:
  - Each cell is the probability of that combination, $f_{X,Y}(x, y)$
- Probability that we randomly select a woman who favors gay marriage?

$$f_{X,Y}(1, 1) = \mathbb{P}(X = 1, Y = 1) = 0.3$$

- Probability that we randomly select a man who favors gay marriage?

$$f_{X,Y}(0, 1) = \mathbb{P}(X = 0, Y = 1) = 0.22$$

Marginal distributions

- Marginal distribution: the probability distribution of one of the r.v.s.
  - (Psst, just what we covered last week)
- Computing marginals from the joint p.m.f.:

$$f_Y(y) = \mathbb{P}(Y = y) = \sum_x f_{X,Y}(x, y)$$

- Intuition: sum over the probability that $Y = y$ for all possible values of x
  - Works because these are disjoint events that partition the space of X
  - Law of Total Probability in action!

Example: marginals for gay marriage

	Favor Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.3	0.21	$0.3 + 0.210.51$
Male $X = 0$	0.22	0.27	$0.22 + 0.270.49$
Marginal	$0.3 + 0.220.52$	$0.21 + 0.270.48$	

- What's the $f_Y(1) = \mathbb{P}(Y = 1)$?
  - Probability that a man favors gay marriage plus the probability that a woman favors gay marriage.

$$f_Y(1) = f_{X,Y}(1,1) + f_{X,Y}(0,1) = 0.3 + 0.22 = 0.52$$

  - Disjoint sets $\rightsquigarrow$ add their probabilities!
- Works for all marginal distributions!

Conditional distributions

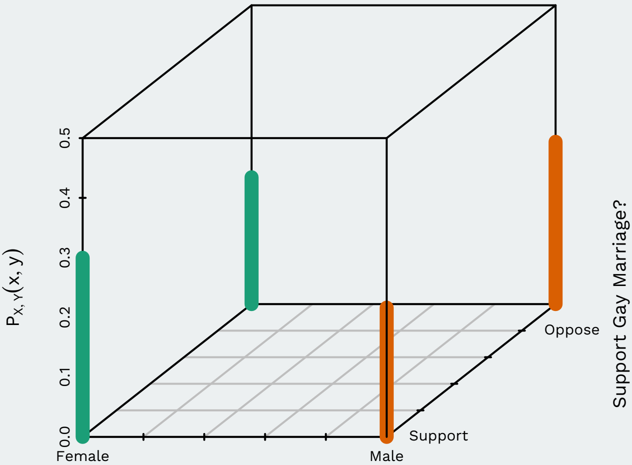
- Conditional distribution: distribution of Y if we know $X = x$.
- **Definition**: The **conditional probability mass function** or conditional p.m.f. of Y conditional of X is

$$f_{Y|X}(y|x) = \frac{\mathbb{P}(\{X = x\} \cap \{Y = y\})}{\mathbb{P}(X = x)} = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

- Intuitive definition:

$$f_{Y|X}(y|x) = \frac{\text{Probability that } X = x \text{ and } Y = y}{\text{Probability that } X = x}$$

Example: conditionals for gay marriage



Example: conditionals for gay marriage

	Favor Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.3	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- Probability of favoring gay marriage conditional on being a man?

$$f_{Y|X}(y = 1|x = 0) = \frac{\mathbb{P}(\{X = 0\} \cap \{Y = 1\})}{\mathbb{P}(X = 0)} = \frac{0.22}{0.22 + 0.27} = 0.44$$

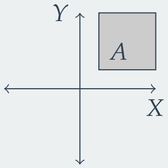
Example: conditionals for gay marriage

	Favor Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.3	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- Probability of favoring gay marriage conditional on being a woman?

$$f_{Y|X}(y = 1|x = 1) = \frac{\mathbb{P}(\{X = 1\} \cap \{Y = 1\})}{\mathbb{P}(X = 1)} = \frac{0.3}{0.3 + 0.21} = 0.59$$

Continuous r.v.s



- Now, let's think about the case where X and Y are continuous.
- $\mathbb{P}(X = x, Y = y) = 0$ for similar reasons as with single r.v.s.
- We will focus on getting the probability of being in some subset of the 2-dimensional plane.

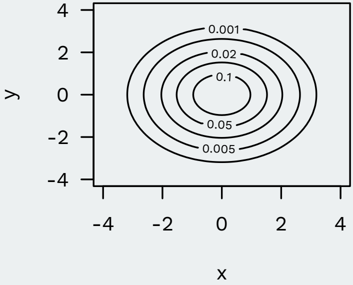
Continuous joint p.d.f.

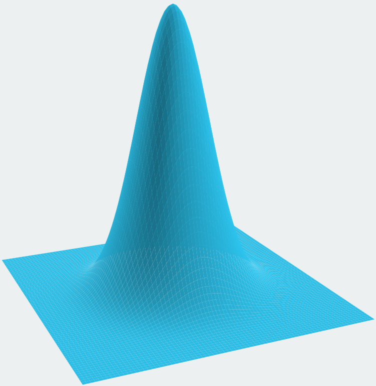
- **Definition**: For two continuous r.v.s X and Y, the **joint probability density function** (or joint p.d.f.) $f_{X,Y}(x,y)$ is a function such that:
  1. $f_{X,Y}(x,y) \geq 0$ for all values of (x,y), (non-negative)
  2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$, (probability of anything is 1)
  3. for any subset A of the xy-plane,

$$\mathbb{P}((X,Y) \in A) = \int_A \int f_{X,Y}(x,y)dxdy$$

- $\mathbb{P}((X,Y) \in A) =$ volume or density under the curve

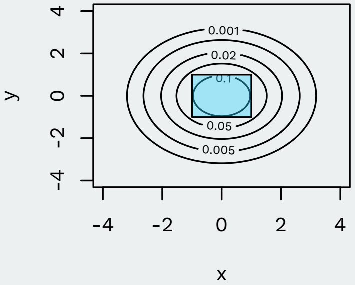
Joint densities are 3D

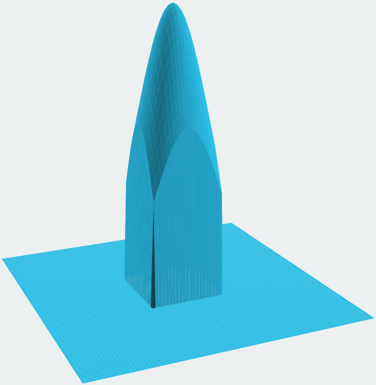




- X and Y axes are on the "floor," height is the value of $f_{X,Y}(x,y)$.
- Remember $f_{X,Y}(x,y) \neq \mathbb{P}(X = x, Y = y)$.

Probability = volume





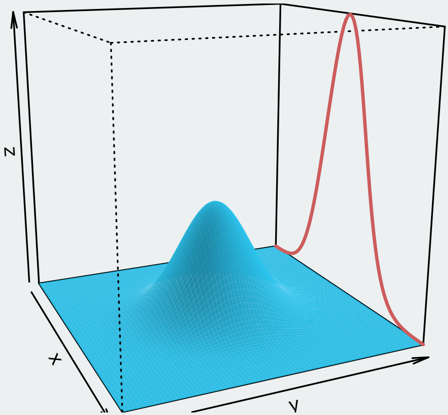
- Probability = volume above a specific region.

Continuous marginal distributions

- We can recover the marginal PDF of one of the variables by integrating over the distribution of the other variable:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

Visualizing continuous marginals



- To get the marginal of X, crush all of the mass of the p.d.f. $f_{X,Y}(x,y)$ across all the values of y.
- Think of the integral like a snow plow in the winter, piling up all of the density into one sliver.

Continuous conditional distributions

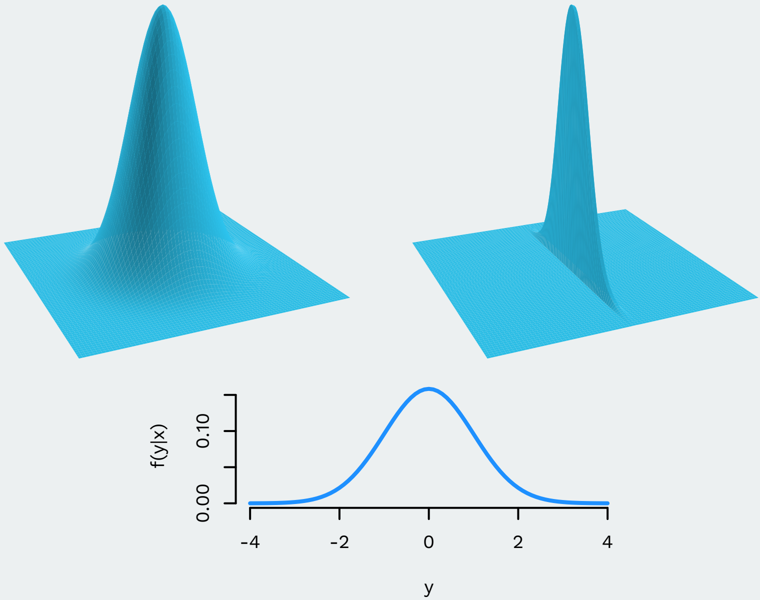
- **Definition**: the conditional pdf of a continuous random variable is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

assuming that $f_X(x) > 0$. Then, we have the following:

$$\mathbb{P}(a < Y < b|X = x) = \int_a^b f_{Y|X}(y|x)dy.$$

Conditional distributions as slices



2/ Properties of Joint Distributions

Properties of joint distributions

- Single r.v.s had a center and a spread.
- With 2 r.v.s, we can additionally measure how strong the dependence is between the variables.
- These will be the crucial building blocks for assessing relationships between variables in real data.

Expectations over multiple r.v.s

- For multiple r.v.s, we have to take expectations over the joint distribution.
- With discrete X and Y:

$$\mathbb{E}[g(X,Y)] = \sum_x \sum_y g(x,y)\, f_{X,Y}(x,y)$$

- Example: suppose we wanted to known the expectation of XY:

$$\mathbb{E}[XY] = \sum_x \sum_y xy\, f_{X,Y}(x,y)$$

Joint expectation example

	Favor Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.3	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- What's the expectation $\mathbb{E}[XY]$?

$$
\begin{aligned}
\mathbb{E}[XY] &= \sum_x \sum_y xy\, f_{X,Y}(x,y) \\
&= 1 \cdot 1 \cdot f_{X,Y}(1,1) + 1 \cdot 0 \cdot f_{X,Y}(1,0) \\
&\quad + 0 \cdot 1 \cdot f_{X,Y}(0,1) + 0 \cdot 0 \cdot f_{X,Y}(0,0) \\
&= 1 \cdot 1 \cdot f_{X,Y}(1,1) = 0.3
\end{aligned}
$$

Independence

- **Definition**: two r.v.s Y and X are independent (which we write $X \perp\!\!\!\perp Y$) if for all sets A and B:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

- Knowing the value of X gives us no information about the value of Y.
- If X and Y are independent, then:
  - $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ (joint is the product of marginals)
  - $f_{Y|X}(y|x) = f_Y(y)$ (conditional is the marginal)
  - $h(X) \perp\!\!\!\perp g(Y)$ for any functions $h()$ and $g()$ (functions of independent r.v.s are independent)

Key properties of independent r.v.s

- **Theorem** If X and Y are independent r.v.s, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

- Proof for discrete X and Y:

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_x \sum_y xy\, f_{X,Y}(x,y) \\
&= \sum_x \sum_y xy\, f_X(x) f_Y(y) \\
&= \left(\sum_x x\, f_X(x)\right)\left(\sum_y y\, f_Y(y)\right) \\
&= \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

Why independence?

- Independence signals the absence of a relationship between two variables.
- Later on, we'll have to make assumptions that two variables are independent.
- Example:
  - $X = 1$ for getting a get-out-the-vote mailer, $X = 0$ no mailer
  - $Y = 1$ if the person voted, $Y = 0$ if they abstained
  - If X is not independent of background covariates, then we might find a spurious relationship between X and Y.

Conditional Independence

- **Definition**: Two r.v.s X and Y are **conditionally independent** given Z (written $X \perp\!\!\!\perp Y|Z$) if

$$f_{X,Y|Z}(x,y|z) = f_{X|Z}(x|z)f_{Y|Z}(y|z).$$

- X and Y are independent within levels of Z.
- Example:
  - X = swimming accidents, Y = number of ice cream cones sold.
  - In general, dependent.
  - Conditional on Z = temperature, independent.

Covariance

- If two variables are not independent, how do we measure the strength of their dependence?
  - Covariance
  - Correlation
- Covariance: how do two r.v.s vary together?
  - How often do high values of X occur with high values of Y?

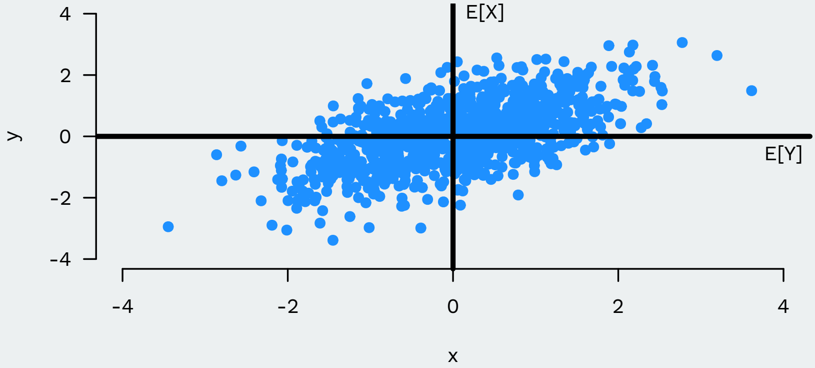
Defining covariance

- **Definition**: The covariance between two r.v.s, X and Y is defined as:

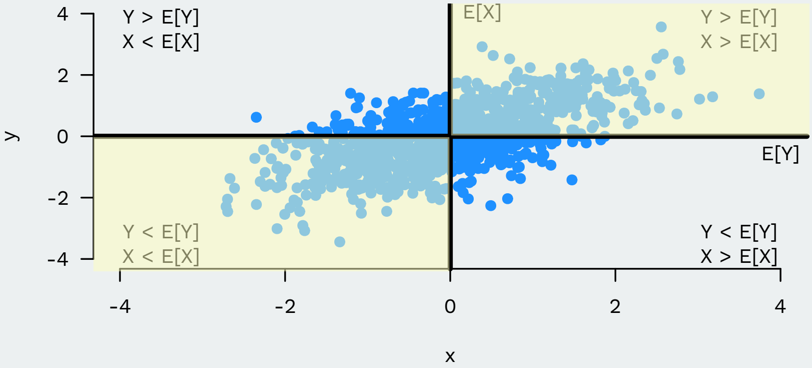
$$\text{Cov}[X, Y] = \mathbb{E}\Big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\Big]$$

- We can show that $\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

Covariance intuition

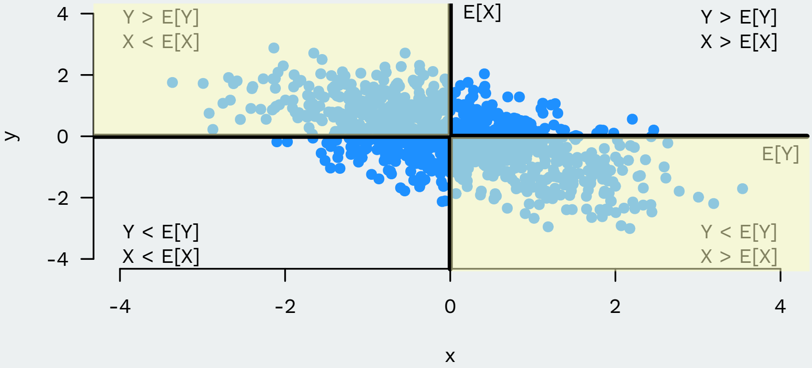


Covariance intuition



- Large values of X tend to occur with large values of Y:
  - $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = (\text{pos. num.})(\text{pos. num}) = +$
- Small values of X tend to occur with small values of Y:
  - $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = (\text{neg. num.})(\text{neg. num}) = +$
- If these dominate $\rightsquigarrow$ positive covariance.

Covariance intuition



- Large values of X tend to occur with small values of Y:
  - $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = (\text{pos. num.})(\text{neg. num}) = -$
- Small values of X tend to occur with large values of Y:
  - $(X - \mathbb{E}[X])(Y - \mathbb{E}[Y]) = (\text{neg. num.})(\text{pos. num}) = -$
- If these dominate $\rightsquigarrow$ negative covariance.

Independence implies zero covariance

- What should $\mathrm{Cov}[X, Y]$ be when $X \perp\!\!\!\perp Y$? Zero!
- Why?

$$\begin{aligned}\mathrm{Cov}[X, Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[X]\mathbb{E}[Y] = 0\end{aligned}$$

- Restated: independent r.v.s have 0 covariance.

Zero covariance doesn't imply independence

- Does $\mathrm{Cov}[X, Y] = 0$ imply that $X \perp\!\!\!\perp Y$? **No!**
- **Counterexample:** $X \in \{-1, 0, 1\}$ with equal probability and $Y = X^2$ (see notes).
- Covariance is a measure of **linear dependence**, so it can miss non-linear dependence.

Properties of variances and covariances

- Properties of covariances:

1. $\text{Cov}[aX + b, cY + d] = ac\text{Cov}[X, Y]$.
2. $\text{Cov}[X, X] = \mathbb{V}[X]$

- Properties of variances that we can state now that we know covariance:

1. $\mathbb{V}[aX + bY + c] = a^2\mathbb{V}[X] + b^2\mathbb{V}[Y] + 2ab\text{Cov}[X, Y]$
2. If X and Y independent, $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$.

Example: intercoder reliability

- Suppose we have two RAs coding documents as positive to negative on a 100-point scale.
- X_1 is the score for coder 1, X_2 for coder 2.
- Both coders are equally precise: $\mathbb{V}[X_1] = \mathbb{V}[X_2] = 4$.
- Usual strategy: take the average coding of the two: $\frac{(X_1+X_2)}{2}$
- Question: when will we get the best (lowest variance) measures of negativity?
  - When coders are independent so that $\text{Cov}[X_1, X_2] = 0$?
  - When coders tend to agree so that $\text{Cov}[X_1, X_2] = 1$?
  - When coders tend to disagree so that $\text{Cov}[X_1, X_2] = -1$?

Independent coders

- What is the variance of the average coding if the codings are independent?

$$\begin{aligned}\mathbb{V}\left[\frac{(X_1 + X_2)}{2}\right] &= \frac{1}{4}\mathbb{V}[X_1] + \frac{1}{4}\mathbb{V}[X_2] \\ &= \frac{1}{4} \cdot 4 + \frac{1}{4} \cdot 4 \\ &= 1 + 1 \\ &= 2\end{aligned}$$

Agreeing coders

- What if they tend to agree so that $\mathrm{Cov}(X_1, X_2) = 1$.

$$
\begin{aligned}
\mathbb{V}\left[\frac{(X_1 + X_2)}{2}\right] &= \frac{1}{4}\mathbb{V}[X_1] + \frac{1}{4}\mathbb{V}[X_2] + 2\frac{1}{2}\frac{1}{2}\mathrm{Cov}[X_1, X_2] \\
&= \frac{1}{4} \cdot 4 + \frac{1}{4} \cdot 4 + \frac{1}{2} \cdot 1 \\
&= 1 + 1 + \frac{1}{2} \\
&= 2.5
\end{aligned}
$$

Disagreeing coders

- What if they tend to disagree so that $\text{Cov}(X_1, X_2) = -1$.

$$\begin{aligned}
\mathbb{V}\left[\frac{(X_1 + X_2)}{2}\right] &= \frac{1}{4}\mathbb{V}[X_1] + \frac{1}{4}\mathbb{V}[X_2] + 2\frac{1}{2}\frac{1}{2}\text{Cov}[X_1, X_2] \\
&= \frac{1}{4}\cdot 4 + \frac{1}{4}\cdot 4 + \frac{1}{2}\cdot -1 \\
&= 1 + 1 - \frac{1}{2} \\
&= 1.5
\end{aligned}$$

Correlation

- Covariance is not scale-free: $\text{Cov}[2X, Y] = 2\text{Cov}[X, Y]$
  - $\rightsquigarrow$ hard to compare covriances across different r.v.s
  - Is a relationship stronger? Or just do to rescaling?

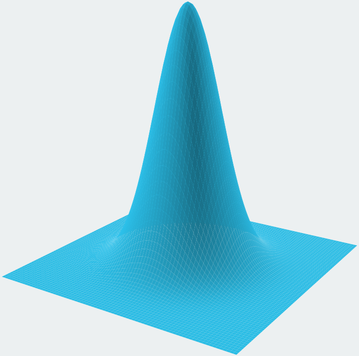
- Correlation is a scale-free measure of linear dependence.
- **Definition**: The **correlation** between two r.v.s X and Y is defined as:

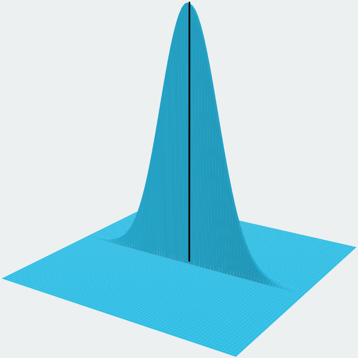
$$\rho = \rho(X, Y) = \frac{\text{Cov}[X, Y]}{\sqrt{\mathbb{V}[X]\mathbb{V}[Y]}} = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y}$$

- Basic idea: covariance dividing out the scales of the respective variables.
- Correlation properties:
  - $-1 \leq \rho \leq 1$
  - if $|\rho(X, Y)| = 1$, then X and Y are perfectly correlated with a deterministic linear relationship: $Y = a + bX$.

3/ Conditional Expectation

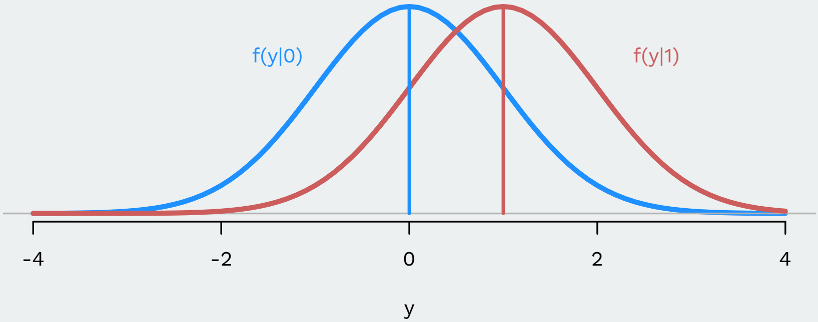
Why conditional expectations?





- With univariate distributions, we summarized them with the expectation and the variance.
- Conditional distributions are also univariate distribution and so we can summarize them with its mean and variance.

Conditional expectations are important



- Gives us insight into a key question:
  - How does the mean of Y change as we change X?
- Examples:
  - Expected number of coups across different types of political institution.
  - Expectated ideology for at different income levels.

Defining condition expectations

- **Definition**: The **conditional expectation** of Y conditional on $X = x$ is:

$$\mathbb{E}[Y|X = x] = \begin{cases} \displaystyle\sum_y y\, f_{Y|X}(y|x) & \text{discrete } Y \\ \displaystyle\int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)dy & \text{continuous } Y \end{cases}$$

- Intuition: exactly the same definition of the expected value with $f_{Y|X}(y|x)$ in place of $f_Y(y)$
- The expected value of the (univariate) conditional distribution.

Calculating conditional expectations

	Favor Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.3	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	

- What's the conditional expectation of support for gay marriage Y given someone is a man $X = 0$?

$$
\begin{aligned}
\mathbb{E}[Y|X=0] &= \sum_y y\, f_{Y|X}(y|x=0) \\
&= 0 \times f(y=0|x=0) + 1 \times f(y=1|x=0) \\
&= 1 \times \frac{0.22}{0.22+0.27} \\
&= 0.44
\end{aligned}
$$

Conditional expectations are random variables

- For a particular x, we can calculate $\mathbb{E}[Y|X = x]$.
- But X takes on many possible values with uncertainty
  $\rightsquigarrow$ $\mathbb{E}[Y|X]$ takes on many possible values with uncertainty.
- $\rightsquigarrow$ Conditional expectations are random variables!
- Binary X:

$$\mathbb{E}[Y|X] = \begin{cases} \mathbb{E}[Y|X = 0] & \text{with prob. } \mathbb{P}(X = 0) \\ \mathbb{E}[Y|X = 1] & \text{with prob. } \mathbb{P}(X = 1) \end{cases}$$

- Has an expectation, $\mathbb{E}\big[\mathbb{E}[Y|X]\big]$, and a variance, $\mathbb{V}\big[\mathbb{E}[Y|X]\big]$.

Law of iterated expectations

- Average/mean of the conditional expectations: $\mathbb{E}\big[\mathbb{E}[Y|X]\big]$.
    - Can we connect this to the marginal (overall) expectation?
- **Theorem** (The Law of Iterated Expectations): If the expectation exist and for discrete X,

$$\mathbb{E}[Y] = \mathbb{E}\left[\mathbb{E}[Y|X]\right] = \sum_{x} \mathbb{E}[Y|X = x] f_X(x)$$

Example: law of iterated expectations

	Favor Gay Marriage $Y = 1$	Oppose Gay Marriage $Y = 0$	Marginal
Female $X = 1$	0.3	0.21	0.51
Male $X = 0$	0.22	0.27	0.49
Marginal	0.52	0.48	1

- $\mathbb{E}[Y|X = 1] = 0.59$ and $\mathbb{E}[Y|X = 0] = 0.44$.
- $f_X(1) = 0.51$ (females) and $f_X(0) = 0.49$ (males).
- Plug into the iterated expectations:

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[Y|X]] &= \mathbb{E}[Y|X = 0]f_X(0) + \mathbb{E}[Y|X = 1]f_X(1) \\
&= 0.44 \times 0.49 + 0.59 \times 0.51 \\
&= 0.52 = \mathbb{E}[Y]
\end{aligned}$$

Properties of conditional expectations

1. $\mathbb{E}[c(X)|X] = c(X)$ for any function $c(X)$.
   - Example: $\mathbb{E}[X^2|X] = X^2$ (If we know X, then we also know X^2)
2. If X and Y are independent r.v.s, then
$$\mathbb{E}[Y|X = x] = \mathbb{E}[Y].$$
3. If $X \perp\!\!\!\perp Y|Z$, then
$$\mathbb{E}[Y|X = x, Z = z] = \mathbb{E}[Y|Z = z].$$

Conditional Variance

- Conditional variance describes the spread of the conditional distribution around the conditional expectation.
- **Definition**: The **conditional variance** of a discrete Y given $X = x$ is defined as:

$$\mathbb{V}[Y|X = x] = \sum_y (y - \mathbb{E}[Y|X = x])^2 f_{Y|X}(y|x)$$

- Remember the formula for variance: $\mathbb{V}[Y] = \sum_y (y - \mathbb{E}[Y])^2 f_Y(y)$
- Same idea, replacing marginals with conditionals.

Conditional variance is a random variable

- Again, $\mathbb{V}[Y|X]$ is a random variable and a function of X, just like $\mathbb{E}[Y|X]$. With a binary X:

$$\mathbb{V}[Y|X] = \begin{cases} \mathbb{V}[Y|X=0] & \text{with prob. } \mathbb{P}(X=0) \\ \mathbb{V}[Y|X=1] & \text{with prob. } \mathbb{P}(X=1) \end{cases}$$

Law of total variance

- We can also relate the marginal variance to the conditional variance and the conditional expectation.
- **Theorem** (Law of Total Variance/EVE's law):

$$\mathbb{V}[Y] = \mathbb{E}[\mathbb{V}[Y|X]] + \mathbb{V}[\mathbb{E}[Y|X]]$$

- The total variance can be decomposed into:
  1. the average of the within group variance ($\mathbb{E}[\mathbb{V}[Y|X]]$) and
  2. how much the average varies between groups ($\mathbb{V}[\mathbb{E}[Y|X]]$).

4/ Sums and Means of Random Variables

Sums and means are random variables

- If X_1 and X_2 are r.v.s, then $X_1 + X_2$ is a r.v.
  - Has a mean $\mathbb{E}[X_1 + X_2]$ and a variance $\mathbb{V}[X_1 + X_2]$
- The **sample mean** is a function of sums and so it is a r.v. too:

$$\bar{X} = \frac{X_1 + X_2}{2}$$

Distribution of sums/means

	X_1	X_2	$X_1 + X_2$	$\bar{X}$
draw 1	80	83	163	81.5
draw 2	42	47	89	44.5
draw 3	49	45	94	47
draw 4	37	2	39	19.5
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
			distribution of the sum	distribution of the mean

Independent and identical r.v.s

- We often will work with independent and identically distributed r.v.s, $X_1, \dots, X_n$
  - Random sample of n respondents on a survey question.
  - Written "i.i.d."
- **Independent**: $X_i \perp\!\!\!\perp X_j$ for all $i \neq j$
- **Identically distributed**: $f_{X_i}(x)$ is the same for all i
  - $\mathbb{E}[X_i] = \mu$ for all i
  - $\mathbb{V}[X_i] = \sigma^2$ for all i

Note on calculating expectations

- A lot of what we have seen so far has been calculating $\mathbb{E}[X_i]$ or $\mathbb{E}[Y|X = x]$ by plugging into the definition:

$$\mathbb{E}[X_i] = \sum_x x\, f_{X_i}(x)$$

- Moving forward, we will often do the following:
  - Represent the mean of a particular variable in terms of a parameter: $\mathbb{E}[X_i] = \mu$
  - Calculate the mean of a function of variables in terms of that parameter.
- The expected values will come from known properties of particular families of distributions.

Coders

- Going back to the idea of having coders.
- Imagine that $X_1, \dots, X_n$ are n i.i.d. negativity scores from different coders.
- The true level of negativity is μ and each coder:
  - Get the right level of negativity on average: $\mathbb{E}[X_i] = \mu$
  - Has some spread around the right answer: $\mathbb{V}[X_i] = \sigma^2$
- We might want to know if:
  1. Does using the sample mean of these coders also give us the right answer on average?
  2. Does using the sample mean reduce the noise of the coders?

Sample sums/mean of i.i.d. r.v.s

- What is the expectation of the sample mean of the i.i.d. r.v.s?

$$
\begin{aligned}
\mathbb{E}[\bar{X}_n] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] \\
&= \frac{1}{n}n\mu \\
&= \mu
\end{aligned}
$$

- The expectation of the sample mean is just the mean of each observation.
- Sample mean $\rightsquigarrow$ right answer on average.

Variance of the sample mean

- What about the variance of the sample mean of i.i.d. r.v.s?

$$
\begin{aligned}
\mathbb{V}[\bar{X}_n] &= \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathbb{V}[X_i] \\
&= \frac{1}{2} n\sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

- Variance of the sample mean is the variance of each observation divided by the number of observations.
- More coders $\rightsquigarrow$ lower variance.

Law of Large Numbers

- **Theorem** (Weak Law of Large Numbers) Let $X_1, \ldots, X_n$ be a an i.i.d. draws from a distribution with mean $E[X_i] = \mu$. Then, as n gets large, the distribution of $\bar{X}_n$ will collapse to a spike on μ.
- Technically, $\bar{Y}_n$ converges in probability to μ, which means for all (small) values $\varepsilon > 0$:

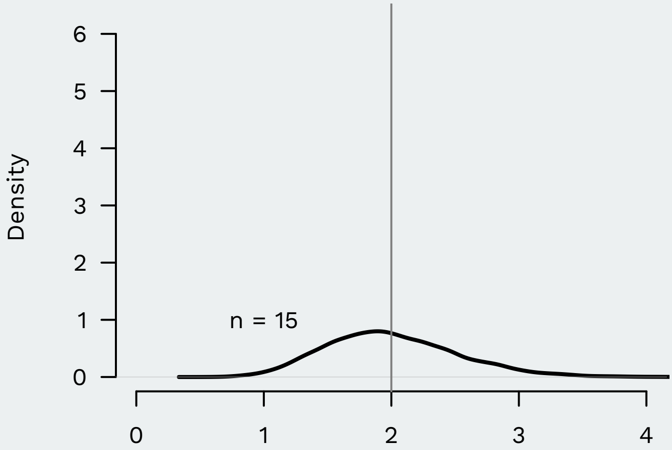
$$\lim_{n \to \infty} \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) = 0$$

- Intuition: The probability of $\bar{X}_n$ being "far away" from μ goes to 0 as n gets big.

LLN by simulation in R

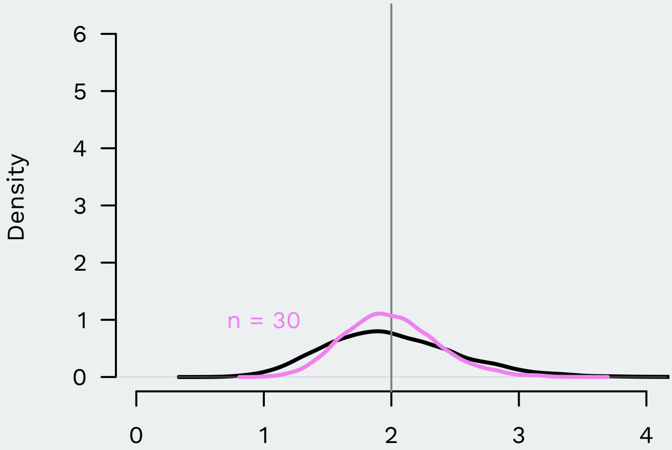
```
nsims <- 10000
holder <- matrix(NA, nrow = nsims, ncol = 6)
bad.holder <- matrix(NA, nrow = nsims, ncol = 6)
for (i in 1:nsims) {
    s5 <- rexp(n = 5, rate = 0.5)
    s15 <- rexp(n = 15, rate = 0.5)
    s30 <- rexp(n = 30, rate = 0.5)
    s100 <- rexp(n = 100, rate = 0.5)
    s1000 <- rexp(n = 1000, rate = 0.5)
    s10000 <- rexp(n = 10000, rate = 0.5)

    holder[i, 1] <- mean(s5)
    holder[i, 2] <- mean(s15)
    holder[i, 3] <- mean(s30)
    holder[i, 4] <- mean(s100)
    holder[i, 5] <- mean(s1000)
    holder[i, 6] <- mean(s10000)
}
```

LLN in action



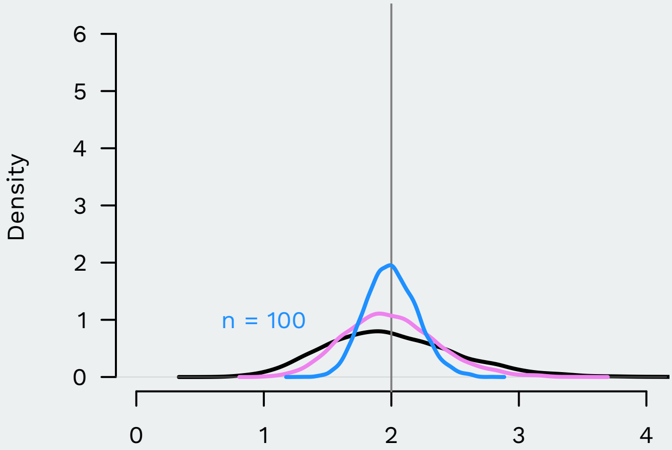
- Distribution of $\bar{X}_{15}$

LLN in action



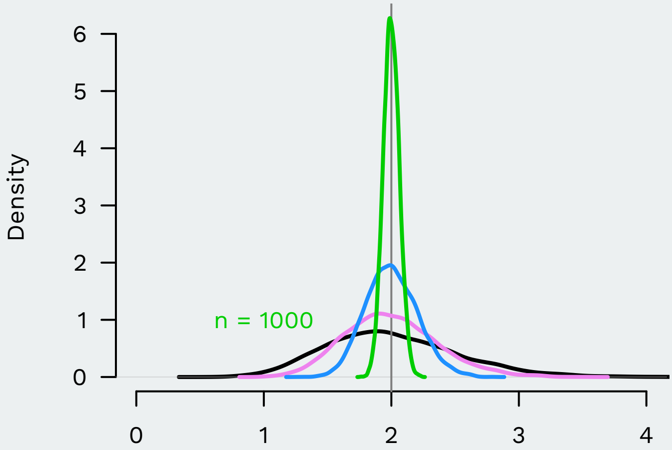
- Distribution of $\bar{X}_{30}$

LLN in action



- Distribution of $\bar{X}_{100}$

LLN in action



- Distribution of $\bar{X}_{1000}$

5/ Wrap-up

Review

- Multiple r.v.s require joint p.m.f.s and joint p.d.f.s
- Multiple r.v.s can have distributions that exhibit dependence as measured by covariance and correlation.
- The conditional expectation of one variable on the other is an important quantity that we'll see over and over again.
- The sample mean of a series of random variables is also a random variable and has a distribution.