Multivariate Distributions

Ye Wang
University of North Carolina at Chapel Hill

Mathematics and Statistics For Political Research
POLI783

Multivariate distributions

- It is rare that we are interested in only one random variable.
- Many questions of interest in social science are about the relationship between multiple r.v.s.
- E.g., are voters with a higher level of education (X) increasingly supportive of the Democratic Party (Y)?
- Does tax cut (X) result in faster economic growth (Y)?
- Are Get-Out-To-Vote (GOTV) campaigns (X) more effective in promoting the turnout rate (Y) among senior voters (Z)?

Joint distributions

- For two r.v.s, X and Y, we define their joint c.d.f. as

$$F_{X,Y}(x,y) = \mathbb{P}(X \leq x, Y \leq y).$$

- If both X and Y are discrete, we can also define their joint p.m.f. as

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y).$$

- Clearly, $p_{X,Y}(x,y) \geq 0$ and $\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{X,Y}(x,y) = 1$.

Joint distributions

	Support ($Y = 1$)	Oppose ($Y = 0$)
Female ($X = 1$)	0.32	0.19
Male ($X = 0$)	0.29	0.20

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

$$p_{X,Y}(1,1) = \mathbb{P}(X = 1,\ Y = 1) = 0.32.$$

Joint distributions

- For continuous r.v.s X and Y, we can define their joint p.d.f as

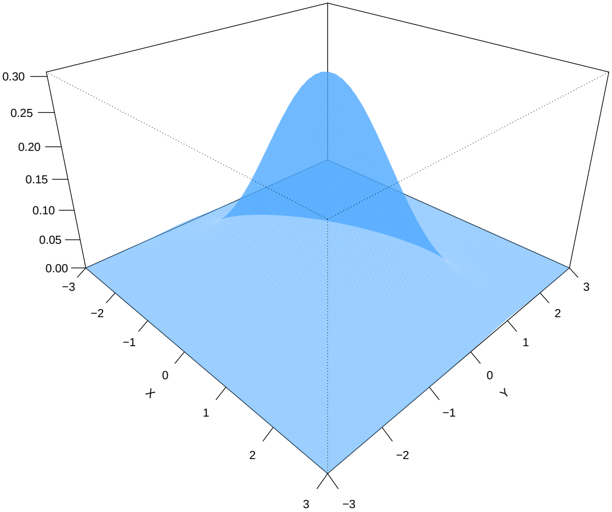
$$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}.$$

- Using the Newton-Leibniz formula, we know that

$$\mathbb{P}((X,Y) \in A) = \iint_{(x,y)\in A} f_{X,Y}(x,y)\, dx\, dy.$$

- E.g., if X is income and Y is ideology, what is the proportion of moderate conservatives whose income is below 2k per month?
- We can see that $f_{X,Y}(x,y) \geq 0$, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx\, dy = 1$, and $\mathbb{P}(X = x, Y = y) = 0$.

Joint distributions



Expectations over multiple r.v.s

- 2-d LOTUS: take expectations over the joint distribution.
- With discrete X and Y:

$$\mathbb{E}[g(X,Y)] = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} g(x,y)p_{X,Y}(x,y).$$

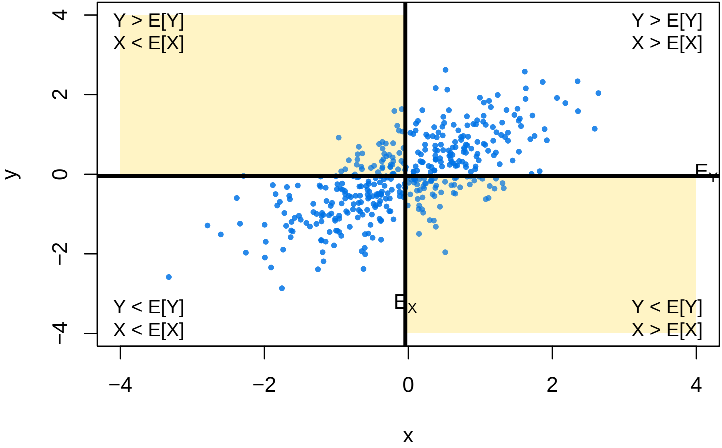
- With continuous X and Y:

$$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dx\,dy.$$

Covariance

- To measure the strength of the dependence between two r.v.s, we can calculate their covariance:

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



Covariance

- From the definition of covariance, we can verify that

$$
\begin{aligned}
&\mathrm{Cov}[X, X] = \mathrm{Var}[X],\\
&\mathrm{Cov}[X, c] = 0,\\
&\mathrm{Cov}[aX_1 + bX_2, Y] = a\,\mathrm{Cov}[X_1, Y] + b\,\mathrm{Cov}[X_2, Y],\\
&\mathrm{Var}[X + Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\,\mathrm{Cov}[X, Y],\\
&\mathrm{Var}\left[\sum_{k=1}^{K} X_k\right] = \sum_{k=1}^{K} \mathrm{Var}[X_k] + 2\sum_{k=1}^{K}\sum_{k'<k} \mathrm{Cov}[X_k, X_{k'}].
\end{aligned}
$$

Covariance

- If X and Y are independent, then $\mathrm{Cov}[X, Y] = 0$.
- The converse is not true!
- Consider $U \sim \mathrm{Uniform}(-1, 1)$ and define: $X = U$, $Y = U^2$.
- $\mathbb{E}[X] = 0$, and

$$\mathbb{E}[Y] = \int_{-1}^{1} u^2 \cdot \tfrac{1}{2} du = \tfrac{1}{3},$$
$$\mathbb{E}[XY] = \int_{-1}^{1} u^3 \cdot \tfrac{1}{2} du = 0.$$

- Therefore, $\mathrm{Cov}[X, Y] = 0$ even though $Y = X^2$.
- A zero covariance implies that X and Y are not linearly dependent.

Correlation

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

$$\begin{aligned}\rho = &\rho(X, Y) = \frac{\mathrm{Cov}[X, Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} \\ &= \mathrm{Cov}\left(\frac{X - \mathbb{E}[X]}{\sigma_X}, \frac{Y - \mathbb{E}[Y]}{\sigma_Y}\right).\end{aligned}$$

- Covariance after dividing out the scales of the respective variables.
- $-1 \le \rho \le 1$, and $|\rho(X, Y)| = 1$ if and only if X and Y are perfectly correlated with a deterministic linear relationship: $Y = a + bX$.

Covariance matrix

- For a multivariate r.v. $\mathbf{X} = (X_1, X_2, \ldots, X_K)$, its expectation is a vector

$$\mathbb{E}\left[\mathbf{X}\right] = \left(\mathbb{E}\left[X_1\right], \mathbb{E}\left[X_2\right], \ldots, \mathbb{E}\left[X_K\right]\right)'.$$

and its variance is a matrix

$$\begin{aligned} \mathrm{Var}\left[\mathbf{X}\right] =& \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])'\right] \\ =& \begin{pmatrix} \mathrm{Var}\left[X_1\right] & \mathrm{Cov}\left[X_1, X_2\right] & \ldots & \mathrm{Cov}\left[X_1, X_K\right] \\ \mathrm{Cov}\left[X_1, X_2\right] & \mathrm{Var}\left[X_2\right] & \ldots & \mathrm{Cov}\left[X_2, X_K\right] \\ \ldots & \ldots & \ddots & \ldots \\ \mathrm{Cov}\left[X_K, X_1\right] & \mathrm{Cov}\left[X_K, X_2\right] & \ldots & \mathrm{Var}\left[X_K\right] \end{pmatrix}, \end{aligned}$$

which is known as the covariance matrix.

Multivariate standard normal distribution

- If $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_K)'$ consists of i.i.d. r.v.s that follow $\mathcal{N}(0,1)$, their joint p.d.f. is

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{K/2}} \exp\left(-\frac{\mathbf{z}'\mathbf{z}}{2}\right),$$

  where $\mathbf{z} = (z_1, z_2, \ldots, z_K)'$.
- Easy to see the mean/variance: $\mathbb{E}[\mathbf{Z}] = 0$ and $\text{Var}[\mathbf{Z}] = \mathbf{I}_K$.
- A multivariate general normal distribution (MVN) is denoted as $\mathcal{N}(\mu, \Sigma)$.
- If (Z_1, Z_2, Z_3) follow a MVN, then (Z_1, Z_2) also follow a MVN.
- If (Z_1, Z_2) follow a MVN and $\text{Cov}\left[Z_1, Z_2\right] = 0$, then Z_1 and Z_2 are independent.

Marginal distributions

- The joint distribution depicts the randomness in both X and Y.
- We can derive the distribution of either variable by marginalizing over the other variable.
- For discrete r.v.s X and Y, we have

$$\mathbb{P}(Y = y) = \sum_{x \in \mathcal{X}} p_{X,Y}(x, y).$$

- We sum over $p_{X,Y}(x, y)$ for all possible values of X.
- This is known as the marginal p.m.f. of Y.
- Marginal expectation:

$$\mathbb{E}[Y] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} y p_{X,Y}(x, y) = \sum_{y \in \mathcal{Y}} y p_Y(y).$$

Marginal distributions

	Support ($Y = 1$)	Oppose ($Y = 0$)	**Marginal**
Female ($X = 1$)	0.32	0.19	0.51
Male ($X = 0$)	0.29	0.20	0.49
Marginal	0.61	0.39	

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

$$\begin{aligned}\mathbb{P}(Y = 1) &= \mathbb{P}(X = 1, Y = 1) + \mathbb{P}(X = 0, Y = 1) \\ &= 0.32 + 0.29 = 0.61.\end{aligned}$$

- Works for all marginals.

Marginal distributions

- For continuous r.v.s X and Y, we can derive Y's marginal distribution by integrating over X:

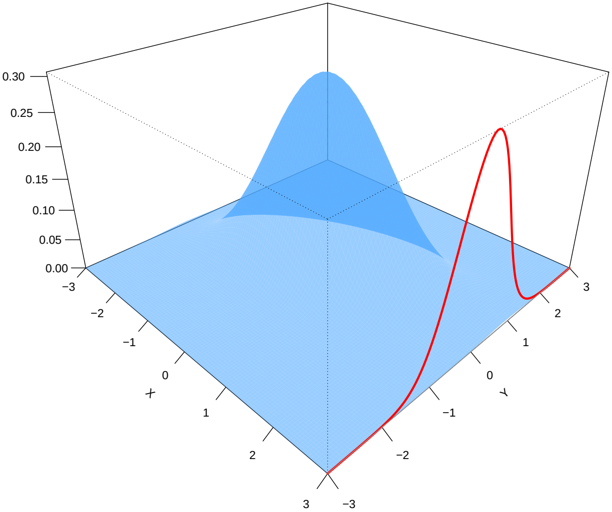
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx.$$

- Similarly, to derive X's marginal distribution, we integrate over Y:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy.$$

- Marginalization projects the density function to a single dimension.
- Knowing the joint distribution of income and ideology, we can derive the distribution of each variable.
- The converse is not true as marginal distributions do not contain information on their interaction.
- It can be generalized to distributions with more r.v.s.

Marginal distributions



Linearity of expectation

- With the concept of marginal distributions, we can prove the linearity of expectation:

$$
\begin{aligned}
\mathbb{E}[aX + bY] = &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(aX + bY)f_{X,Y}(x,y)dxdy \\
=&a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}Xf_{X,Y}(x,y)dxdy + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}Yf_{X,Y}(x,y)dxdy \\
=&a\int_{-\infty}^{\infty}X\left(\int_{-\infty}^{\infty}f_{X,Y}(x,y)dy\right)dx \\
&+ b\int_{-\infty}^{\infty}Y\left(\int_{-\infty}^{\infty}f_{X,Y}(x,y)dx\right)dy \\
=&a\int_{-\infty}^{\infty}Xf_X(x)dx + b\int_{-\infty}^{\infty}Yf_Y(y)dy \\
=&a\mathbb{E}[X] + b\mathbb{E}[Y].
\end{aligned}
$$

Conditional distributions

- We can use conditional distributions to describe the conditional probability for a r.v. to take different values.
- For discrete r.v.s X and Y, we can define the conditional p.m.f. of Y given $X = x$:

$$\mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)},$$

  where $\mathbb{P}(X = x) \neq 0$.
- We can verify that $\mathbb{P}(Y = y \mid X = x) \geq 0$ and $\sum_{y \in \mathcal{Y}} \mathbb{P}(Y = y \mid X = x) = 1$.

Conditional distributions

	Support ($Y = 1$)	Oppose ($Y = 0$)	**Marginal**
Female ($X = 1$)	0.32	0.19	0.51
Male ($X = 0$)	0.29	0.20	0.49
Marginal	0.61	0.39	

- Probability of favoring gay marriage conditional on male?
- $\mathbb{P}(Y = 1 \mid X = 0) = \frac{\mathbb{P}(Y=1,X=0)}{\mathbb{P}(X=0)} = \frac{0.29}{0.29+0.2} \approx 0.592.$

Bayes' rule for r.v.s

- We can generalize the Bayes' rule to r.v.s. using conditional p.m.f.s
- First, the law of total probability can be expressed as

$$\mathbb{P}(X = x) = \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x \mid Y = y)\mathbb{P}(Y = y).$$

- Next, the Bayes' rule:

$$\mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(X = x \mid Y = y)\mathbb{P}(Y = y)}{\mathbb{P}(X = x)}.$$

Conditional distributions

- For continuous r.v.s X and Y, we can define the conditional p.d.f. of Y given X:

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

for all values of X where $f_X(x) > 0$.

- Therefore,

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

- If two continuous r.v.s, X and Y, are independent, then

$$f_{X,Y}(X = x, Y = y) = f_X(x) f_Y(y),$$
$$f_{Y|X}(y \mid x) = f_Y(y).$$

Conditional distributions

- Let's prove that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ if X and Y are independent:

$$\begin{aligned} \mathbb{E}[XY] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XYf_{X,Y}(x,y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} XYf_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty} Xf_X(x)\left(\int_{-\infty}^{\infty} Yf_Y(y)dy\right)dx \\ &= \int_{-\infty}^{\infty} Xf_X(x)\left(\mathbb{E}[Y]\right)dx = \mathbb{E}[X]\mathbb{E}[Y]. \end{aligned}$$

Conditional expectation

- The conditional expectation of Y is a functional of Y's conditional distribution:

$$\begin{aligned}\mu(\mathbf{x}) =& \mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}] \\ =& \begin{cases} \sum_{y \in \mathcal{Y}} y \mathbb{P}(Y = y \mid \mathbf{X} = \mathbf{x}) & \text{discrete r.v.} \\ \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x) dy & \text{continuous r.v.} \end{cases}\end{aligned}$$

- What is the average of Y if we know that $\mathbf{X} = \mathbf{x}$?
- $\mu(\mathbf{X})$ is a function of $\mathbf{X}$ and known as the conditional expectation function (CEF).
- How does the average of Y change with the value of $\mathbf{X}$?

Conditional expectation

	Support ($Y = 1$)	Oppose ($Y = 0$)	**Marginal**
Female ($X = 1$)	0.32	0.19	0.51
Male ($X = 0$)	0.29	0.20	0.49
Marginal	0.61	0.39	

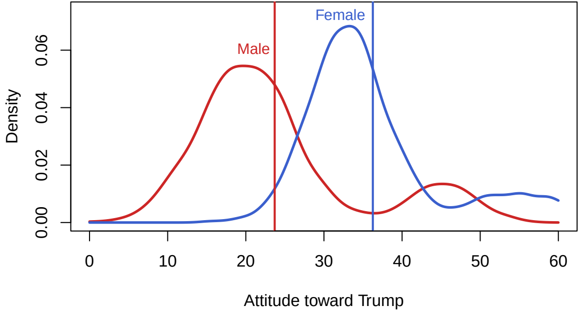
- Conditional expectation of gay marriage support Y among men $X = 0$?

$$\begin{aligned}\mathbb{E}[Y \mid X = 0] =& \sum_y y\mathbb{P}(Y = y \mid X = 0) \\ =& 0 \times \mathbb{P}(Y = 0 \mid X = 0) + 1 \times \mathbb{P}(Y = 1 \mid X = 0) \\ =& 1 \times \frac{0.29}{0.29 + 0.20} \approx 0.592\end{aligned}$$

- Similarly, $\mathbb{E}[Y \mid X = 0] = \frac{0.32}{0.32+0.19} \approx 0.627$.

CEF for discrete covariates

- Suppose Y is one's feeling thermometer toward Mr. Trump.
- $X \in \{\text{White}, \text{Black}, \text{Hispanic}, \text{Asian}, \text{Native Americans}\}$.
- $\mathbb{E}[Y \mid X = x]$ captures the average feeling of people from a specific ethnic group.
- We can further define $Z \in \{\text{Male}, \text{Female}\}$ and $\mathbb{E}[Y \mid X = x, Z = z]$.
- These values show the concentration of different distributions.



CEF for continuous covariates

- When $\mathbf{X}$ include continuous variables, $\mu(\mathbf{X})$ can take infinite values.
- The functional form of $\mu(\mathbf{X})$ is determined by the joint distribution of $\mathbf{X}$ and Y.
- If $(X, Y) \sim \mathcal{N}(\mu, \Sigma)$, then

$$\mu(X) = \mathbb{E}[Y] + \frac{\text{Cov}[X, Y]}{\text{Var}[X]}(X - \mathbb{E}[X]).$$

- The goal of many empirical research is to infer $\mu(\mathbf{X})$ from data.
- E.g., how does the average ideology of individuals vary with their income?
- $\mu(\mathbf{X})$ is also a r.v. as its values are determined by $\mathbf{X}$.
- Its expectation: $\mathbb{E}\left[\mathbb{E}[Y \mid \mathbf{X}]\right]$; its variance: $\text{Var}\left[\mathbb{E}[Y \mid \mathbf{X}]\right]$.

Properties of conditional expectations

- Linearity:

$$\mathbb{E}[aY_1 + bY_2 \mid \mathbf{X}] = a\mathbb{E}[Y_1 \mid \mathbf{X}] + b\mathbb{E}[Y_2 \mid \mathbf{X}].$$

- For any function of $\mathbf{X}$, $g(\mathbf{X})$,

$$\begin{aligned}
&\mathbb{E}\left[g(\mathbf{X}) \mid \mathbf{X}\right] = g(\mathbf{X}), \\
&\mathbb{E}\left[g(\mathbf{X})Y \mid \mathbf{X}\right] = g(\mathbf{X})\mathbb{E}[Y \mid \mathbf{X}] = g(\mathbf{X})\mathbb{E}\left[\mu(\mathbf{X}) \mid \mathbf{X}\right].
\end{aligned}$$

- If Y and $\mathbf{X}$ are independent,

$$\mathbb{E}[Y \mid \mathbf{X}] = \mathbb{E}[Y]$$

- If $Y \perp\!\!\!\perp Z \mid \mathbf{X}$ are independent,

$$\mathbb{E}[Y \mid \mathbf{X}, Z] = \mathbb{E}[Y \mid \mathbf{X}].$$

Law of iterated expectation

- If $\mathbb{E}[Y] < \infty$, then $\mathbb{E}[Y] = \mathbb{E}\left[\mathbb{E}[Y \mid \mathbf{X}]\right]$:

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}[Y \mid \mathbf{X}]\right] &= \int_{-\infty}^{\infty} \mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}] f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} y f_{Y|\mathbf{X}}(y \mid \mathbf{x}) dy \right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&= \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} f_{Y,\mathbf{X}}(y, \mathbf{x}) d\mathbf{x} \right) dy \\
&= \int_{-\infty}^{\infty} y f_{Y}(y) dy
\end{aligned}$$

- For a discrete r.v. X,

$$\begin{aligned}
\mathbb{E}[Y] =& \mathbb{E}[\mu(X)] = \sum_{x \in \mathcal{X}} \mu(X) \mathbb{P}\left[X = x\right] \\
&= \sum_{x \in \mathcal{X}} \mathbb{E}[Y \mid X = x] \mathbb{P}\left[X = x\right].
\end{aligned}$$

- Generally, $\mathbb{E}[Y \mid \mathbf{X}] = \mathbb{E}\left[\mathbb{E}[Y \mid \mathbf{X}, \mathbf{Z}] \mid \mathbf{X}\right]$ (tower property).

Law of iterated expectation

	Support ($Y = 1$)	Oppose ($Y = 0$)	**Marginal**
Female ($X = 1$)	0.32	0.19	0.51
Male ($X = 0$)	0.29	0.20	0.49
Marginal	0.61	0.39	

- $\mathbb{E}[Y \mid X = 1] = 0.627$ and $\mathbb{E}[Y \mid X = 0] = 0.592$.
- $\mathbb{P}[X = 1] = 0.51$ and $\mathbb{P}[X = 0] = 0.49$.
- Plugging into the formula:

$$
\begin{aligned}
\mathbb{E}[Y] = \mathbb{E}\left[\mathbb{E}[Y \mid X]\right] = \sum_{x \in \mathcal{X}} x\mathbb{E}[Y \mid X = x] \\
=&\mathbb{E}[Y \mid X = 1]\mathbb{P}[X = 1] + \mathbb{E}[Y \mid X = 0]\mathbb{P}[X = 0] \\
=&0.627 * 0.51 + 0.592 * 0.49 \approx 0.61.
\end{aligned}
$$

CEF as the best predictor

- $\mathbb{E}[Y \mid \mathbf{X}]$ can be seen as a "projection" of Y on $\mathbf{X}$.
- It captures the variation in Y that can be predicted by $\mathbf{X}$.
- We can prove that no function of $\mathbf{X}$ can predict Y more accurately.
- Let's define $e_\mu = Y - \mu(\mathbf{X})$, the part in Y that cannot be explained by $\mu(\mathbf{X})$.
- e_μ is also known as the error of the CEF, with $\mathbb{E}\left[e_\mu \mid \mathbf{X}\right] = 0$.
- Then, for any function $g(\mathbf{X})$, we have

$$\mathbb{E}\left[e_\mu^2\right] = \mathbb{E}\left[(Y - \mu(\mathbf{X}))^2\right] \leq \mathbb{E}\left[(Y - g(\mathbf{X}))^2\right] = \mathbb{E}\left[e_g^2\right].$$

- This is one way to define the CEF.

```
## glX
##   1
```

CEF as the best predictor

- First note that for any function $h(\mathbf{X})$,

$$
\begin{aligned}
&\mathbb{E}\left[(Y-\mu(\mathbf{X}))h(\mathbf{X})\right]=\mathbb{E}\left[e_{\mu}h(\mathbf{X})\right]\\
=&\mathbb{E}\left[\mathbb{E}\left[(Y-\mu(\mathbf{X}))h(\mathbf{X})\mid\mathbf{X}\right]\right]\\
=&\mathbb{E}\left[h(\mathbf{X})\mathbb{E}\left[Y\mid\mathbf{X}\right]\right]-\mathbb{E}\left[h(\mathbf{X})\mathbb{E}\left[\mu(\mathbf{X})\mid\mathbf{X}\right]\right]\\
=&\mathbb{E}\left[h(\mathbf{X})\mu(\mathbf{X})\right]-\mathbb{E}\left[h(\mathbf{X})\mu(\mathbf{X})\right]=0.
\end{aligned}
$$

- Therefore, we have

$$
\begin{aligned}
&\mathbb{E}\left[(Y-g(\mathbf{X}))^2\right]=\mathbb{E}\left[(Y-\mu(\mathbf{X})+\mu(\mathbf{X})-g(\mathbf{X}))^2\right]\\
=&\mathbb{E}\left[(Y-\mu(\mathbf{X}))^2\right]+2\mathbb{E}\left[(Y-\mu(\mathbf{X}))(\mu(\mathbf{X})-g(\mathbf{X}))\right]\\
&+\mathbb{E}\left[(\mu(\mathbf{X})-g(\mathbf{X}))^2\right]\\
=&\mathbb{E}\left[(Y-\mu(\mathbf{X}))^2\right]+\mathbb{E}\left[(\mu(\mathbf{X})-g(\mathbf{X}))^2\right]\geq\mathbb{E}\left[(Y-\mu(\mathbf{X}))^2\right].
\end{aligned}
$$

Conditional variance

- The conditional variance of a Y given $\mathbf{X} = \mathbf{x}$ is

$$\sigma^2(\mathbf{x}) = \text{Var}[Y \mid \mathbf{X} = \mathbf{x}] = \mathbb{E}\left[(Y - \mu(\mathbf{x}))^2 \mid \mathbf{X} = \mathbf{x}\right].$$

- Spread of the conditional distribution around its expectation.
- By definition, it equals the variance of the CEF error:

$$\text{Var}[Y \mid \mathbf{X} = \mathbf{x}] = \text{Var}[e_\mu \mid \mathbf{X} = \mathbf{x}] = \mathbb{E}[e_\mu^2 \mid \mathbf{X} = \mathbf{x}].$$

- Can be re-expressed in the usual way:

$$\text{Var}[Y \mid \mathbf{X} = \mathbf{x}] = \mathbb{E}\left[Y^2 \mid \mathbf{X} = \mathbf{x}\right] - (\mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}])^2.$$

- Similarly, $\text{Var}[Y \mid \mathbf{X}]$ is a function of $\mathbf{X}$ and a r.v.
- We say e_μ^2 is homoskedastic if $\sigma^2(\mathbf{x}) = \sigma^2$ and heteroskedastic otherwise.

Law of total variance

- Similar to the tower property, we have the following relationship between $\mathrm{Var}[Y]$ and $\mathrm{Var}[Y \mid \mathbf{X}]$:

$$\mathrm{Var}[Y] = \mathbb{E}\left[\mathrm{Var}[Y \mid \mathbf{X}]\right] + \mathrm{Var}\left[\mathbb{E}\left[Y \mid \mathbf{X}\right]\right].$$

- Suppose $\mathbf{X}$ includes only discrete values (e.g., ethnicity), then its value defines disjoint groups.
- $\mathbb{E}\left[Y \mid \mathbf{X}\right]$ is Y's expectation within each group and $\mathrm{Var}[Y \mid \mathbf{X}]$ is Y's variance within each group.
- $\mathbb{E}\left[\mathrm{Var}[Y \mid \mathbf{X}]\right]$: the expected within-group variance; $\mathrm{Var}\left[\mathbb{E}\left[Y \mid \mathbf{X}\right]\right]$: the variance of group-level expectations (between-group variance).

Law of total variance

- Let's prove it:

$$
\begin{aligned}
\mathrm{Var}[Y] = \mathbb{E}\left[(Y - \mathbb{E}[Y])^2\right] &= \mathbb{E}\left[\mathbb{E}\left[(Y - \mathbb{E}[Y])^2 \mid \mathbf{X}\right]\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[Y^2 - 2Y\mathbb{E}[Y] + (\mathbb{E}[Y])^2 \mid \mathbf{X}\right]\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[Y^2 - (\mathbb{E}[Y \mid \mathbf{X}])^2 + (\mathbb{E}[Y \mid \mathbf{X}])^2 - 2Y\mathbb{E}[Y] + (\mathbb{E}[Y])^2 \mid \mathbf{X}\right]\right] \\
=& \mathbb{E}\left[\mathbb{E}\left[Y^2 \mid \mathbf{X}\right] - (\mathbb{E}[Y \mid \mathbf{X}])^2\right] + \mathbb{E}\left[\mathbb{E}\left[(\mathbb{E}[Y \mid \mathbf{X}] - \mathbb{E}[Y])^2 \mid \mathbf{X}\right]\right] \\
=& \mathbb{E}\left[\mathrm{Var}[Y \mid \mathbf{X}]\right] + \mathbb{E}\left[(\mathbb{E}[Y \mid \mathbf{X}] - \mathbb{E}[Y])^2\right] \\
=& \mathbb{E}\left[\mathrm{Var}[Y \mid \mathbf{X}]\right] + \mathrm{Var}\left[\mathbb{E}\left[Y \mid \mathbf{X}\right]\right].
\end{aligned}
$$