$\boxed{02/21}$

3.6.1 Covariance & correlation

"Education & income are correlated"

Def ~~Cov~~ let X, Y be r.v's. Covariance:

$$\mathrm{Cov}(X,Y) := \mathbb{E}\left[(X-\mathbb{E}(X))(Y-\mathbb{E}Y)\right].$$

Correlation:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

← no units.

Prop (Properties) $\forall$ r.v's X, Y:

1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
2. $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$
3. $-1 \le \rho_{X,Y} \le 1$

1 - obvious from def

2 - direct check (do it!)

3 $\iff$ $|\mathrm{Cov}(X,Y)| \le \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}$

$\iff$ $\mathbb{E}\left[(X-\mathbb{E}X)(Y-\mathbb{E}Y)\right] \le \sqrt{\mathbb{E}[(X-\mathbb{E}X)^2]}\sqrt{\mathbb{E}[(Y-\mathbb{E}Y)^2]}$

True by Cauchy-Schwarz.

Ex (X = # bedrooms, Y = # iPods) – p. 65.

$p(x,y)$:

$x \backslash y$	0	1	2
0	0.05	0.12	0.03
1	0.07	0.1	0.08
2	0.02	0.26	0.27

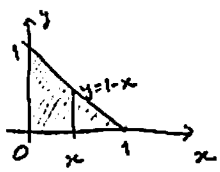
$E[X] = 1.35, \quad E[Y] = 1.24$

$$E[XY] = \sum_{x=0}^{2} \sum_{y=0}^{2} xy\, p(x,y)$$ (discrete version of Prop on p.87)

$$= 1\cdot1\cdot0.1 + 1\cdot2\cdot0.08 + 2\cdot1\cdot0.26 + 2\cdot2\cdot0.27 = 1.86$$

$$\mathrm{Cov}(X,Y) = 1.86 - 1.35 \times 1.24 = 0.186$$ Positive correlation

Ex $(X,Y) \sim \mathrm{Unif}(\Delta)$. $\quad \mathrm{Cov}(X,Y) = ?\ \ \rho_{X,Y} = ?$



$$f(x,y) = \begin{cases} 2, & (x,y) \in \Delta \\ 0, & \text{—} \end{cases}$$

$$E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f(x,y)\,dx\,dy = 2 \iint_\Delta x\,dx\,dy = 2\int_0^1 \left(\int_0^{1-x} x\,dy\right) dx$$

$$= 2\int_0^1 \underbrace{x(1-x)}_{= x - x^2}\,dx = 2\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{1}{3}$$

$E[Y] = \frac{1}{3}$ by symmetry.

$$E[XY] = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\,dx\,dy = 2\iint_\Delta xy\,dx\,dy = 2\int_0^1 x \underbrace{\left(\int_0^{1-x} y\,dy\right)}_{= \frac{y^2}{2}\Big|_0^{1-x} = \frac{(1-x)^2}{2}} dx$$

$$= \int_0^1 \underbrace{x(1-x)^2}_{= x - 2x^2 + x^3}\,dx = 1 - \frac{2}{3} + \frac{1}{4} = \frac{1}{12}$$

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{1}{12} - \frac{1}{3}\cdot\frac{1}{3} = -\frac{1}{36}.$$ Negative correlation

- $\rho_{X,Y} = ?$

$$E[X^2] = 2\iint_\Delta x^2\,dx\,dy = 2\int_0^1 \left(\int_0^{1-x} x^2\,dy\right)dx = 2\int_0^1 \underbrace{x^2(1-x)}_{= x^2 - x^3} = 2\left(\frac{1}{3} - \frac{1}{4}\right) = \frac{1}{6}$$

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}.$$

$\mathrm{Var}(Y) = \frac{1}{18}$ by symmetry.

$$\rho_{X,Y} = \frac{-1/36}{\sqrt{\frac{1}{18}\cdot\frac{1}{18}}} = -\frac{1}{2}.$$ Negative correlation, but weaker than when $Y = -X$.

Linear dependence:

Ex (Strongest correlation). Let $Y = kX + b$. $\quad Cov(X,Y) = ?\quad \rho_{X,Y} = ?$

$E[X] =: \mu_X, \qquad E[Y] = k\mu_X + b.$

- $cov(X,Y) = E[(X-\mu_X)(kX+\cancel{b}-k\mu_X-\cancel{b})] = k\cdot E[(X-\mu_X)^2] = k\cdot Var(X)$

$Var(X) =: \sigma_X, \qquad Var(Y) = k^2\sigma_X$

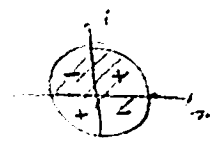
- $\rho_{X,Y} = \dfrac{k\cdot Var(x)}{\sqrt{\sigma_X\cdot k^2\sigma_X}} = \dfrac{k}{|k|} = \begin{cases} 1 & \text{if } k>0 \\ -1 & \text{if } k<0 \end{cases}$ ← strongest possible correlation.

Prop (Weakest corr) If X, Y are independent then $Cov(X,Y) = 0$, ie X, Y are uncorrelated.

Proof: $Cov(X,Y) = \underbrace{E[XY]}_{=E[X]\cdot E[Y]} - E[X]E[Y] = 0$. QED

Ex (Uncorrelated does not always imply independent)

Let $(X,Y) \sim Unif(C)$, $\quad C = \{(x,y) : x^2+y^2 \le 1\}$



Then $E[X] = E[Y] = 0$, $\quad E[XY] = \iint_C xy\,dx\,dy = 0$ (Ex. compute, or argue by symmetry)

$\Rightarrow Cov(X,Y) = 0 - 0 = 0.$

However, X, Y are dependent (as we know).

Summary:
- Correlation $\rho_{X,Y}$ is always between $-1, 1$.
- $\rho_{X,Y} = \pm 1$ if there is a linear dependence between X, Y
  ~~(one can show that this is the only case where $\rho_{X,Y} = \pm 1$)~~
- $\rho_{X,Y} > 0$ means larger values of X tend to imply larger values of Y
  < 0 — — — — — — — smaller — — —
- $\rho_{X,Y} = 0$: "uncorrelated". Happens when X, Y are independent. But not always vice versa.

Further Properties of covariance

~~**Prop**~~ $\cancel{\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)}$

$\cancel{\mathrm{Cov}(aX,Y) = a\,\mathrm{Cov}(X,Y)}, \quad \cancel{\mathrm{Cov}(X,bY) = b\,\mathrm{Cov}(X,Y)}$

~~(This follows easily from def)~~

Prop $\boxed{\mathrm{Cov}\left(\sum_i X_i, \sum_j Y_j\right) = \sum_i \sum_j \mathrm{Cov}(X_i, Y_j)}$ ← (Analog of the numeric identity $\left(\sum_i x_i\right)\left(\sum_j y_j\right) = \sum_i \sum_j x_i y_j$).

Proof $\mu_i := E(X_i), \quad \nu_j := E(Y_j)$

$$\mathrm{Cov}\left(\sum_i X_i, \sum_j Y_j\right) = E\left[\left(\sum_i X_i\right)\left(\sum_j Y_j\right)\right] - E\left[\sum_i X_i\right] E\left[\sum_j Y_j\right] =: I - II$$

$$I = E\left[\sum_i \sum_j X_i Y_j\right] = \sum_i \sum_j E(X_i Y_j) \quad \text{(by linearity of expectation)}$$

$$II = \left(\sum_i E(X_i)\right)\left(\sum_j E(Y_j)\right) = \sum_i \sum_j E(X_i)\, E(Y_j)$$

$$\Rightarrow I - II = \sum_i \sum_j \left(E(X_i Y_j) - E(X_i)E(Y_j)\right) = \sum_i \sum_j \mathrm{Cov}(X_i, Y_j) \qquad \text{QED.}$$

Variance of a sum.

$\mathrm{Var}\left(\sum_i X_i\right) \neq \sum_i \mathrm{Var}(X_i)$ in general (for example, if $Y = -X$, $\mathrm{Var}(X+Y) = 0$ but $\mathrm{Var}(X) + \mathrm{Var}(-X) > 0$.

Instead,

$$\mathrm{Var}\left(\sum_i X_i\right) = \mathrm{Cov}\left(\sum_i X_i, \sum_j X_j\right) = \sum_i \sum_j \mathrm{Cov}(X_i, X_j) \qquad \text{(by Prop. above)}$$

$$= \sum_i \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$$

We have proved:

Cor $\mathrm{Var}\left(\sum_i X_i\right) = \sum_i \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$.

In particular, if X_i are <u>uncorrelated</u> then

$$\mathrm{Var}\left(\sum_i X_i\right) = \sum_i \mathrm{Var}(X_i)$$

Ex X = amount of pain-relieving medicine,

Y = level of pain after taking the medicine

Model:

- $Y = \underset{\substack{\uparrow \\ \text{initial level} \\ \text{of pain}}}{b} - \underbrace{mX}_{\text{effect of medicine}} + \underbrace{Z}_{\text{random fluctuations, indep. of } X}$ ("regression of Y on X")
- $X \sim N(\mu_X, \sigma_X^2)$; $Z \sim N(\mu_Z, \sigma_Z^2)$ independent.

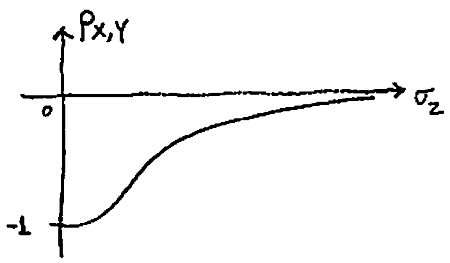
Compute $\rho_{X,Y}$.

$$\mathrm{Cov}(X,Y) = \underbrace{\mathrm{Cov}(X,b)}_{\overset{\shortparallel}{0}\ (b=\text{const})} - m\underbrace{\mathrm{Cov}(X,X)}_{\overset{\shortparallel}{\mathrm{Var}(X)}} + \underbrace{\mathrm{Cov}(X,Z)}_{\overset{\shortparallel}{0}\ \text{by independence}}$$

$$= -m\,\mathrm{Var}(X)$$

$$\mathrm{Var}(Y) = \mathrm{Var}(-mX+Z) = m^2\mathrm{Var}(X) + \mathrm{Var}(Z)$$

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\,\sigma_Y} = \frac{-m\,\sigma_X^2}{\sigma_X\sqrt{m^2\sigma_X^2+\sigma_Z^2}} = -\frac{1}{\sqrt{1+\left(\frac{\sigma_Z}{m\sigma_X}\right)^2}}.$$



$\sigma_Z = 0$ (no fluctuations) $\Rightarrow \rho_{X,Y} = -1$ (strongest)

$\sigma_Z \to \infty$ (overwhelming fluctuations) $\Rightarrow \rho_{X,Y} = 0$ (no pattern).

3.7.2. Random vectors

~~Obs~~ View r.v's $X_1, \dots, X_n$ as a random vector
$$X = (X_1, \dots, X_n) \in \mathbb{R}^n.$$

· Def $\mathbb{E}[X] := (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$ ←

$$\mathrm{Cov}(X) := \mathbb{E}\left[(X - \mathbb{E}X)(X - \mathbb{E}X)^T\right]$$ ← analog of Variance

~~Remark~~ Prop $\mathrm{Cov}(X) = \left[\mathrm{Cov}(X_i, X_j)\right]_{i,j=1}^{n}$

is a symmetric, ~~n×n matrix~~. positive-semidefinite, $n \times n$ matrix

~~Elementwise~~ (ij)-th entry is $\mathbb{E}\left[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)^T\right] = \mathrm{Cov}(X_i, X_j)$.

· Symmetry ~~is obvious:~~ ~~$\mathrm{Cov}(X)^T = \mathbb{E}[$~~
$$\left((X - \mathbb{E}X)(X - \mathbb{E}X)^T\right)^T = (X - \mathbb{E}X)(X - \mathbb{E}X)^T \Rightarrow \text{D}.$$

· PSD: $x^T \mathrm{Cov}(X)\, x = \mathbb{E}\left[\underbrace{x^T(X - \mathbb{E}X)}\,\underbrace{(X - \mathbb{E}X)^T x}\right]$
$$= \mathbb{E}\left[\langle X - \mathbb{E}X, x\rangle^2\right] \geq 0 \quad \forall x \in \mathbb{R}^n.$$

Ex (from p. ...) : $\bar{X} = (X, Y)$.
$$\mathrm{Cov}(\bar{X}) = \begin{bmatrix} \frac{1}{18} & -\frac{1}{36} \\ -\frac{1}{36} & \frac{1}{18} \end{bmatrix}$$

Ex (from p. ...): $\bar{X} = (X, Y)$
$$\mathrm{Cov}(\bar{X}) = \begin{bmatrix} \sigma_X^2 & -m\sigma_X^2 \\ -m\sigma_X^2 & m^2\sigma_X^2 + \sigma_Z^2 \end{bmatrix}$$

Def R. vector Z is *isotropic* if $\mathbb{E}[Z]=0$, $\mathrm{Cov}(Z)=I_n$ — identity $\begin{bmatrix}1 & & 0\\ & \ddots & \\ 0 & & 1\end{bmatrix}$

Prop $Z=(Z_1,\dots,Z_n)$: Z_i indep. r.v's with $\mathbb{E}Z_i=0$, $\mathbb{E}Z_i^2=1$.
Then Z is isotropic

Γ $\mathrm{Cov}(Z) = \mathbb{E}ZZ^T \Rightarrow \mathrm{cov}(Z)_{ij} = \mathbb{E}Z_iZ_j = \begin{cases}1, & i=j\\ 0, & i\neq j.\end{cases}$ $\rfloor$

Prop Suppose Z is isotropic. Then

$$X := \mu + AZ$$

satisfies

$$\mathbb{E}[X]=\mu,\quad \mathrm{Cov}(X)=AA^T.$$

Γ • $\mathbb{E}[X] = \mu + \mathbb{E}[AZ] = \mu + A\underbrace{\mathbb{E}[Z]}_{=0} = \mu.$

• (for $\mu=0$) : $\mathrm{Cov}(X) = \mathbb{E}[XX^T] = \mathbb{E}[(AZ)(AZ)^T]$
$= \mathbb{E}[AZZ^TA^T] = A\underbrace{\mathbb{E}(ZZ^T)}_{=I}A^T = AA^T$ $\rfloor$

Remark (PSD matrices) Recall from linear algebra:
~~$\forall A$: $n\times n$~~ (i) $\forall$ $n\times n$ matrix A, $\Sigma := AA^T$ is PSD (Check!)
(ii) $\forall$ PSD Σ $\exists$ symmetric A: $\Sigma = AA^T = A^2$
$A =: \sqrt{\Sigma}$.

Cor $\forall$ PSD Σ, $\exists$ r.v. X with $\mathrm{cov}(X)=\Sigma$.