Chapter 19

Expected Values, Covariance and Correlation

Lecture 19 | Devore 9e: 5.2

19.1 Expectation of a function of two variables

Theorem 19.1. *For $h : \mathbb{R}^2 \to \mathbb{R}$,*

$$\mathbb{E}[h(X,Y)] = \begin{cases} \displaystyle\sum_x \sum_y h(x,y)\, p(x,y), & \textit{discrete}, \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f(x,y)\, \mathrm{d}x\, \mathrm{d}y, & \textit{continuous}. \end{cases}$$

Example 19.2 (Seats)**.** Five seats in a row are numbered $1, \dots, 5$. You and one other person sit in two distinct seats chosen at random; X is your seat, Y theirs. So

$$p(x,y) = \begin{cases} \frac{1}{20}, & x, y \in \{1, \dots, 5\},\ x \neq y, \\ 0, & \text{else}, \end{cases}$$

since there are $5 \times 4 = 20$ equally likely ordered pairs. Let $h(x,y) = |x-y| - 1$ be the number of seats *between* you. Then

$$\mathbb{E}[h(X,Y)] = \sum_x \sum_y h(x,y) p(x,y) = \frac{1}{20} \sum_{x \neq y} \big(|x-y| - 1\big) = \frac{20}{20} = 1.$$

The sum is easiest to organize by the gap $d = |x-y|$: there are $2(5-d)$ ordered pairs with gap d, each contributing $d-1$, so

$$\sum_{x \neq y} \big(|x-y| - 1\big) = \sum_{d=1}^{4} 2(5-d)(d-1) = 8 \cdot 0 + 6 \cdot 1 + 4 \cdot 2 + 2 \cdot 3 = 20.$$

Example 19.3 (Cost of a can of nuts)**.** Almonds cost \$1.50/lb, cashews \$2.25/lb, peanuts \$0.75/lb. With X, Y the almond and cashew weights and $1 - X - Y$ the peanut weight, the cost of a can is

$$h(X,Y) = 1.5X + 2.25Y + 0.75(1 - X - Y) = 0.75 + 0.75X + 1.5Y.$$

With $f(x,y) = 24xy$ on the triangle,

$$\mathbb{E}[h(X,Y)] = \int_0^1 \int_0^{1-x} \big(0.75 + 0.75x + 1.5y\big)\, 24xy \,\mathrm{d}y\,\mathrm{d}x = \$1.65.$$

Remark 19.4 (Shortcut)**.** Because h is *linear*, the double integral was avoidable. Using the marginal means $\mathbb{E}[X] = \mathbb{E}[Y] = \frac{2}{5}$ computed later this lecture,

$$\mathbb{E}[h(X,Y)] = 0.75 + 0.75\,\mathbb{E}[X] + 1.5\,\mathbb{E}[Y] = 0.75 + 0.30 + 0.60 = \$1.65,$$

and note that this needed *no* assumption of independence — expectation is linear regardless. Doing the integral once in full is still worth it, so that you trust the shortcut when you use it later.

19.2 Conditional expectation

Definition 19.5 (Conditional expectation)**.**

$$\mathbb{E}[h(X) \mid Y = y] = \int_{-\infty}^{\infty} h(x)\, f_{X|Y}(x \mid y)\,\mathrm{d}x.$$

Remark 19.6. This is a *function of* y, not a number: for each value of Y you get a different average. Writing $\mathbb{E}[h(X) \mid Y]$ (with a capital Y) therefore denotes a random variable.

Theorem 19.7 (Law of total expectation / tower property)**.**

$$\mathbb{E}\big[\ \mathbb{E}[h(X) \mid Y]\ \big] = \mathbb{E}[h(X)].$$

Proof. The $f_Y(y)$ cancels, and then the order of integration is swapped:

$$\mathbb{E}\big[\mathbb{E}[h(X) \mid Y]\big] = \int \bigg(\int h(x)\frac{f(x,y)}{f_Y(y)}\,\mathrm{d}x\bigg) f_Y(y)\,\mathrm{d}y = \iint h(x) f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int h(x)\bigg(\int f(x,y)\,\mathrm{d}y\bigg)\mathrm{d}x = \int h(x) f$$

$\square$

This is the expectation version of the law of total probability: average within each case, then average the case averages.

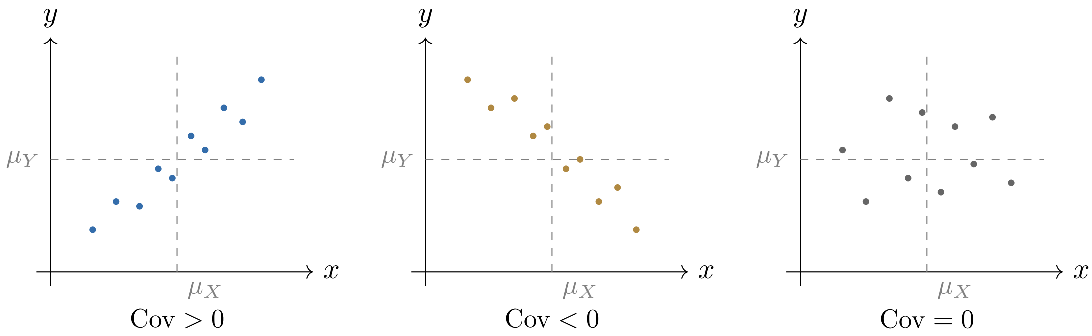
19.3 Covariance

Variance measured how far X strays from its own mean. Covariance measures whether X and Y stray *together*.

Definition 19.8 (Covariance)**.**

$$\mathrm{Cov}(X,Y) := \mathbb{E}\big[(X-\mu_X)(Y-\mu_Y)\big] = \begin{cases} \displaystyle\sum_x \sum_y (x-\mu_X)(y-\mu_Y)p(x,y), & \text{discrete,} \\ \displaystyle\iint (x-\mu_X)(y-\mu_Y) f(x,y)\,\mathrm{d}x\,\mathrm{d}y, & \text{continuous.} \end{cases}$$

Remark 19.9 (Intuition)**.** The product $(X-\mu_X)(Y-\mu_Y)$ is positive when both variables are on the same side of their means and negative when they are on opposite sides. Averaging picks out which happens more often.



Theorem 19.10 (Properties)**.** *1.* $\mathrm{Cov}(X,X)=\mathrm{Var}(X)$*;*

2. $\mathrm{Cov}(X,Y)=\mathbb{E}[XY]-\mu_X\mu_Y$ *(the computational formula);*

3. $\mathrm{Cov}(aX+b,\, cY+d)=ac\,\mathrm{Cov}(X,Y)$*;*

4. $X \perp\!\!\!\perp Y \implies \mathrm{Cov}(X,Y)=0$*.* ***The converse is false.***

Property (2) gives the recipe: from the joint pdf get $\mathbb{E}[XY]$ by a double integral; from the marginals get μ_X and μ_Y; subtract.

$$f(x,y) \longrightarrow \mathbb{E}[XY], \qquad f(x,y) \longrightarrow f_X, f_Y \longrightarrow \mu_X, \mu_Y.$$

Example 19.11 (Can of nuts)**.** $f(x,y)=24xy$ on $\{x,y\ge 0,\ x+y\le 1\}$.

$$\mathbb{E}[XY]=\int_0^1\int_0^{1-x} xy\cdot 24xy\,\mathrm{d}y\,\mathrm{d}x = 8\int_0^1 x^2(1-x)^3\,\mathrm{d}x=\frac{2}{15},$$

$$f_X(x)=12x(1-x)^2, \quad f_Y(y)=12y(1-y)^2 \implies \mu_X=\mu_Y=\int_0^1 12x^2(1-x)^2\,\mathrm{d}x = 12\cdot\frac{1}{30}=\frac{2}{5}.$$

Therefore

$$\mathrm{Cov}(X,Y)=\frac{2}{15}-\frac{2}{5}\cdot\frac{2}{5}=\frac{10}{75}-\frac{12}{75}=\boxed{-\frac{2}{75}}.$$

Negative, as it must be: the can weighs one pound, so more almonds means fewer cashews.

19.4 Correlation

Covariance has units — change pounds to grams and it changes by a factor of 454^2. The strength of a relationship should not depend on the choice of units.

Definition 19.12 (Correlation coefficient)**.**

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

Theorem 19.13 (Properties)**.** *1. $\operatorname{Corr}(aX+b, cY+d) = \operatorname{Corr}(X,Y)$ when $a, c > 0$ — units do not matter;*

2. $-1 \le \rho_{X,Y} \le 1$;

3. $\rho_{X,Y} = \pm 1$ if and only if $Y = aX + b$ exactly, for some $a \neq 0$;

4. $X \perp\!\!\!\perp Y \Rightarrow \rho = 0$ ("uncorrelated"), but not conversely.

Example 19.14 (Property 3 in action — the slope is invisible)**.** Let $Y = 100X + 2$. Then $\operatorname{Cov}(X,Y) = 100\operatorname{Var}(X)$ by property 3 of covariance, and $\sigma_Y = 100\,\sigma_X$, so

$$\rho_{X,Y} = \frac{100\operatorname{Var}(X)}{\sigma_X \cdot 100\,\sigma_X} = \frac{100\sigma_X^2}{100\sigma_X^2} = 1.$$

Now let $Y = -0.00001X + 10{,}000$. The slope is minuscule and the intercept enormous, yet by the same cancellation ($\sigma_Y = 0.00001\,\sigma_X$)

$$\rho_{X,Y} = \frac{-0.00001\operatorname{Var}(X)}{\sigma_X \cdot 0.00001\,\sigma_X} = -1.$$

Correlation records only *whether* the relationship is exactly linear and in which direction — never how steep it is. A nearly flat line and a nearly vertical one both give $|\rho| = 1$.

Remark 19.15. The last expression in the definition should look familiar: it is

$$\cos\theta = \frac{\langle u, v\rangle}{\|u\|\,\|v\|}$$

with covariance as the inner product. Correlation is the cosine of the angle between the centered variables — which is exactly why it lands in $[-1, 1]$ and why ± 1 means perfect alignment.

Example 19.16 (Uncorrelated but dependent)**.** Let (X,Y) take the four values $(-4,1), (4,-1), (2,2), (-2,-2)$, each with probability $\frac{1}{4}$. Then $\mu_X = \mu_Y = 0$ and

$$\mathbb{E}[XY] = \tfrac{1}{4}\big(-4 + (-4) + 4 + 4\big) = 0,$$

so $\operatorname{Cov}(X,Y) = 0$ and $\rho = 0$. But X and Y are wildly dependent: knowing $X = -4$ tells you $Y = 1$ exactly.

Remark 19.17 (The one-sentence summary)**.** ρ measures the strength of a **linear** relationship only. A perfect non-linear relationship can have $\rho = 0$. So "uncorrelated" is strictly weaker than "independent", and $\rho = 0$ is never evidence of independence.

Remark 19.18 (Looking ahead)**.** Writing $Y = aX + b + \varepsilon$ with ε a noise term and asking for the a, b that fit best is **linear regression**. The answer turns out to be $a = \rho\,\sigma_Y/\sigma_X$, which is the sense in which ρ "is" the slope.