10) Covariance and Correlation

10.1 Expectation and joint distributions:

Theorem 10.1: Let X and Y be random variables. Let $h: \mathbb{R}^2 \to \mathbb{R}$ be a function so that $h(X, Y)$ is a new random variable.

If X and Y are discrete then:

$$\boxed{E[h(X,Y)] = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} h(x,y) \, p_{XY}(x,y)}$$

If X and Y are jointly continuous with density function f_{XY}, then

$$\boxed{E[h(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) \, f_{XY}(x,y) \, dy \, dx}$$

Example 9.1: (continued) Use thm 10.1 with $h(x,y) = x \cdot y$ to calculate $E[S.N]$ as follows:

$$E[S.N] = \sum_{s=0}^{1} \sum_{n=0}^{2} sn\, p_{SN}(s,n)$$

$$= \frac{1}{4} + 2 \cdot \frac{1}{4} = \frac{3}{4}$$

Similarly calculating $E[S+N]$ using $h(x,y) = x+y$

$$E[S+N] = \sum_{s=0}^{1} \sum_{n=0}^{2} (s+n)\, p_{SN}(s,n)$$

$$= 1 \cdot p_{SN}(0,1) + 1 \cdot p_{SN}(1,0) + 2 \cdot p_{SN}(0,2) +$$

$$2 \cdot p_{SN}(1,1) + 3 \cdot p_{SN}(1,2)$$

$$= \frac{1}{4} \cdot 0 + 2 \cdot 0 + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} = \frac{3}{2}$$

<u>Note</u>: In previous example

$$E[S] = \sum_{s=0}^{1} s\, p_S(s) = \frac{1}{2}$$

$$E[N] = \sum_{n=0}^{2} n\, p_N(n) = \frac{1}{2} + 2 \cdot \frac{1}{4} + 1$$

so that $E[S] + E[N] = 3/2 = E[S+N]$.
Not a coincidence.

<u>Theorem 10.2</u>: (<u>Linearity of expectations</u>):

Let X and Y be random variables. Let $r, s, t \in \mathbb{R}$.
Then

$$\boxed{E[rX + sY + t] = rE[X] + sE[Y] + t}$$

<u>proof</u>: Let us first prove case where X and Y are discrete.

$$E[rX + sY + t] = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} (rx + sy + t) \cdot p_{XY}(x, y)$$

by Thm 10.1

$$= r \sum_{x \in X(\Omega)} x \sum_{y \in Y(\Omega)} \underbrace{\left(p_{XY}(x,y) \right)}_{= p_X(x)} +$$

$$s \sum_{y \in Y(\Omega)} y \underbrace{\left(\sum_{x \in X(\Omega)} p_{XY}(x,y) \right)}_{= p_Y(y)}$$

$$+ \; t \underbrace{\sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} p_{XY}(x,y)}_{= 1}$$

Using Thm 9.3 and property (jm2) and using the definition Def 7.1,

$$E[rX + sY + t] = r \sum_{x \in X(\Omega)} x \, p_X(x) + s \sum_{y \in Y(\Omega)} y \, p_Y(y)$$

$$+ \; t$$

$$= rE[X] + rE[Y] + t$$

Now giving proof when X and Y are jointly continuous:

$$E[rX+sY+t] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(rx+sy+t)\,p_{XY}(x,y)\,dy\,dx$$

by Thm 10.1

$$= r\int_{-\infty}^{\infty} x\Bigg(\underbrace{\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy}_{=f_X(x)}\Bigg)dx \;+$$

$$s\int_{-\infty}^{\infty} y\Bigg(\underbrace{\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx}_{=f_Y(y)}\Bigg)dy$$

$$+\,t\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$$

Using Thm 9.9 and (jd 2) to perform of of the integrals and using defn of expectations Def 7.8 gives:

$$E[rX+sY+t] = r\int_{-\infty}^{\infty} x f_X(x)\,dx + s\int_{-\infty}^{\infty} y f_Y(y)\,dy + t$$

$$= rE[X] + sE[Y] + t$$

$\blacksquare$

Example 10.3: Let $Y_1 \dots Y_n$ be a sequence of independant Bernoulli trials, each with probability of success p, i.e.

$$Y_i \sim Ber(p)$$

Then

$$X = \sum_{k=1}^{n} Y_k \qquad X \sim Bin(n,p)$$

is equal to total number of successes in n trials!

As we discussed in Example 4.12, $X \sim Bin(n,p)$

Thm 10.2 allows to calculate $E[X]$ as

$$E[X] = E\left[\sum_{k=1}^{n} Y_k\right]$$

$$= \sum_{k=1}^{n} E[Y_k] = \sum_{k=1}^{n} p = np.$$

We used expectation of $Ber(p)$ as p.

10.2 Covariance

Linearity does not hold for variance:

$$Var(X+Y) = E[(X+Y-E[X+Y])^2] \quad \text{by Def 7.18}$$

$$= E[(X-E[X]+Y-E[Y])^2] \quad \text{By Thm 10.2}$$

$$= E[(X-E[X])^2 + (Y-E[Y])^2] + 2E[(X-E[X])(Y-E[Y])]$$

$$= Var(X) + Var(Y) + \underbrace{2E[(X-E[X])(Y-E[Y])]}_{\text{covariance}}$$

Defn 10.4: Let X and Y be random variables. The **covariance** between X and Y is defined as

$$\boxed{\mathrm{Cov}[X,Y] = E\big[(X-E[X])(Y-E[Y])\big]}$$

If $\mathrm{Cov}[X,Y]=0$, we say X and Y are **uncorrelated**. Otherwise they are correlated

We see that covariance is positive. Larger X leads us to expect larger Y.

Theorem 10.5: Let X and Y be random variables, let $r,s,t\in\mathbb{R}$. Then

$$\boxed{\mathrm{Var}[rX+sY+t] = r^2\mathrm{Var}[X] + s^2\mathrm{Var}[Y] + 2rs\,\mathrm{Cov}[X,Y]}$$

proof:
$$\mathrm{Var}[rX+sY+t] = E\big[(rX+sY+t-E[rX+sY+t])^2\big]$$
$$= E\big[(rX - rE[X] + sY - sE[Y] + \cancel{t} - \cancel{t})^2\big]$$
$$= E\big[(rX - rE[X] + sY - sE[Y])^2\big]$$
$$= E\big[(r(X-E[X]) + s(Y-E[Y]))^2\big]$$

$$= E\left[r^2(X-E[X])^2 + s^2(Y-E[Y])^2 + 2rs(X-E[X])(Y-E[Y])\right]$$

$$= r^2 E\left[(X-E[X])^2\right] + s^2 E\left[(Y-E[Y])^2\right] + 2rs\, E\left[(X-E[X])(Y-E[Y])\right]$$

$$= r^2 \operatorname{Var}(X) + s^2 \operatorname{Var}(Y) + 2rs \operatorname{Cov}[X,Y]$$

$\blacksquare$

An alternative expression for $\operatorname{Cov}[X,Y]$

<u>Theorem 10.6</u>: Let X and Y be random variables. Then

$$\boxed{\operatorname{Cov}[X,Y] = E[XY] - E[X]E[Y]}$$

<u>proof</u>: $\operatorname{Cov}[(X-E[X])(Y-E[Y])]$ by defn 10.5

$$= E\left[XY - XE[Y] - E[X]Y + E[X]E[Y]\right]$$

$$= E[XY] - E[X].E[Y] - E[X]E[Y] + E[X]E[Y]$$

by Theorem 10.2

$$= E[XY] - E[X]E[Y]$$

$\blacksquare$

Example 9.1 (continued): When the total number of heads N is larger then we have a higher expectation that the second coin lands heads.
Thus we expect S and N to be positively correlated.
Confirming it with a calculation:

$$\mathrm{Cov}[S,N] = E[SN] - E[S]E[N]$$

$$= \frac{3}{4} - \frac{1}{2}\cdot 1 = \frac{1}{4}.$$

Theorem 10.8: If X, Y and Z are random variables and $r,s,t \in \mathbb{R}$, then

$$\boxed{\mathrm{Cov}[rX+sY+t, Z] = r\,\mathrm{Cov}[X,Z] + s\,\mathrm{Cov}[Y,Z]}$$

proof: proof is by calculation using the defn of covariance, linearity of expectation Thm 10.2

$\mathrm{Cov}[rX+sY+t, Z]$

$$= E[(rX+sY+t - E[rX+sY+t])(Z-E[Z])]$$

$$= E[(r(X-E[X]) + s(Y-E[Y]))(Z-E[Z])]$$

$$= r\,E[(X-E[X])(Z-E[Z])] + s\,E[(Y-E[Y])(Z-E[Z])]$$

$$= r\,\mathrm{Cov}[X,Y] + s\,\mathrm{Cov}[Y,Z]$$

$\blacksquare$

Theorem 10.9: If two random variables are independant then their covariance is zero.

$$\boxed{X \perp\!\!\!\perp Y \implies \mathrm{Cov}[X,Y] = 0}$$

proof: First calculate

$$E[XY] = \sum_{x\in X(\Omega)} \sum_{y \in Y(\Omega)} xy\, p_{XY}(x,y)$$ by Thm 10.1

$$= \sum_{x\in X(\Omega)} \sum_{y \in Y(\Omega)} xy\, p_X(x)\, p_Y(y)$$ by independance and Thm 9.12

$$= \sum_{x\in X(\Omega)} x\, p_X(x) \sum_{y \in Y(\Omega)} y\, p_Y(y)$$

$$= E[X]\,E[Y]$$ by defn 7.1

Therefore

$$\text{Cov}[X,Y] = E[XY] - E[X]E[Y]$$
$$= \cancel{E[X]E[Y]} - \cancel{E[X]E[Y]}$$
$$= 0.$$

Proof for jointly continuous random variables X and Y is very similar.

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x,y)\,dx\,dy \quad \text{by Thm 10.1}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_X(x) f_Y(y)\,dx\,dy \quad \text{independance and Thm 9.12}$$

$$= \int_{-\infty}^{\infty} x f_X(x)\,dx \int_{-\infty}^{\infty} y f_Y(y)\,dy$$

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

Hence

$$\text{Cov}[X,Y] = E[XY] - E[X]E[Y] = 0 \quad \blacksquare$$

Note: We established that:

$$\boxed{X \perp\!\!\!\perp Y \Rightarrow E[XY] = E[X]E[Y]}$$

The converse is <u>not</u> true.

<u>Example 10.7</u>: Smarties come in 8 colours: Red, Green, Blue, Yellow, Orange, Brown and Pink.
Denote probability of random smartie being red by p_R.
Similarly for all other colors: $p_G, p_B, p_Y, p_O, p_{BR}, p_V, p_P$

Consider a box with n randomly drawn smarties.
Let Y be <u>number of yellow smarties</u> in box.
Then

$$Y \sim Bin(n, p_Y)$$

Similarly let B be the number of blue smarties in box. Then

$$B \sim Bin(n, p_B)$$

<u>Calculate $Cov[Y, B]$</u>

Solution: According to Thm 10.6,

$$\mathrm{Cov}[Y, B] = E[YB] - E[Y]E[B]$$

As we have calculated the expectation of binomial distribution in Example 10.3 giving as

$$E[Y] = np_Y \quad , \quad E[B] = np_B$$

We still need to calculate $E[YB]$. (*1)

(*1) Method I:
Using Thm 10.1

$$E[YB] = \sum_{x \in X(\Omega)} \sum_{y \in Y(\Omega)} yb \, P_{YB}(y, b)$$

For this we need joint mass function of the Binomial distribution.

$$P_{YB}(y, b) = P(Y = y, B = b)$$

$$= p_Y^y \, p_B^b \, (1 - p_Y - p_B)^{n-y-b} \binom{n}{y+b} \binom{y+b}{b}$$

Binomial factors count the way to choose yellow and blue smarties from all n smarties.

(*1) Method 2: (similar to method in example 10-3):

Introduce indicator random variables:

$$Y_i = \mathbb{1}_{i\text{-th smartie is yellow}} = \begin{cases} 1 & \text{if } i^{th} \text{ smartie is yellow} \\ 0 & \text{otherwise.} \end{cases}$$

$$B_j = \mathbb{1}_{i\text{-th smartie is blue}} = \begin{cases} 1 & \text{if } j^{th} \text{ smartie is blue} \\ 0 & \text{otherwise.} \end{cases}$$

and then

$$Y = \sum_{i=1}^{n} Y_i$$

$$B = \sum_{j=1}^{n} B_j$$

Using this we have

$$\mathrm{Cov}[Y, B] = \mathrm{Cov}\left[\sum_{i=1}^{n} Y_i, \sum_{j=1}^{n} B_j\right]$$

We find

$$\mathrm{Cov}[Y, B] = \mathrm{Cov}\left[\sum_{i=1}^{n} Y_i, B\right]$$

$$= \sum_{i=1}^{n} \mathrm{Cov}[Y_i, B]$$ (using Thm 10.8 repeatedly)

$$= \sum_{i=1}^{n} \mathrm{Cov}[B, Y_i]$$ (Because covariance is symmetric)

$$= \sum_{i=1}^{n} \mathrm{Cov}\left[\sum_{j=1}^{n} B_j, Y_i\right]$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} \mathrm{Cov}[B_j, Y_i]$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} Cov[Y_i, B_i]$$ (Because covariance is symmetric)

Hence

$$Cov[Y, B] = \sum_{i=1}^{n} \sum_{j=1}^{n} Cov[Y_i, B_i]$$

The covariance of indicator random variables is easy to calculate.
We distinguish cases where both refer to the <u>same smartie</u>, i.e. cases where $i=j$ and where they refer to the same smartie and the case where they refer to different smarties.

In the <u>first case</u>: $i=j$

$$Y_i B_j = Y_i B_i = 0 \qquad \text{when } i=j$$

i.e. a smartie <u>can not</u> be both blue and yellow at same time. Thus covariance is

$$Cov[Y_i, B_i] = E[Y_i B_i] - E[Y_i][B_i]$$

$$= 0 - P_Y P_B = -P_Y P_b$$

For the second case: use the fact that one smartie being yellow is independant of another smartie being blue, so

$$Y_i \perp\!\!\!\perp B_j$$

and we use thm 10.9:

Thus

$$\mathrm{Cov}[Y, B] = \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}[Y_i, B_i]$$

$$= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathrm{Cov}[Y_i, B_j] + \sum_{i=1}^{n} \mathrm{Cov}[Y_i, B_i]$$

by Thm 10.9

(due to j being same as i)

$$= \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} 0 + \sum_{i=1}^{n} -p_Y p_b = -n p_y p_b$$

Note: Note how we split up sum over all pairs of indices to where $i \neq j$ and $i = j$

Hence

$$\mathrm{Cov}[Y, B] = -n p_y p_b$$

Note: Note how we split up sum over all pairs of indices to where $i \neq j$ and $i = j$

10.3 The correlation coefficient

The covariance is not a perfect measure of strength of correlation between 2 random variables because it depends on choice of units for random variables. One can however combine the covariance between X and Y with variances of X and Y in such a way to cancel that dependance on choice of units.

Defn 10.10: Let X and Y be random variables. The correlation coefficient $\rho(X,Y)$ is defined as

$$\boxed{\rho(X,Y) = \begin{cases} \dfrac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} & \text{if } \mathrm{Var}(X)\mathrm{Var}(Y) > 0 \\ 0 & \text{otherwise} \end{cases}}$$

The next theorem summarises why correlation coefficient is convenient:
It does <u>not</u> change as you <u>rescale</u> and it is always between -1 and 1.

Theorem 10.11: Let X and Y be random variables and let $r,s,t,u \in \mathbb{R}$. Then

1.
$$\rho(rX+s, tY+u) = \begin{cases} \rho(X,Y) & \text{if } rt > 0 \\ 0 & \text{if } rt = 0 \\ -\rho(X,Y) & \text{if } rt < 0 \end{cases}$$

2.
$$-1 \le \rho(X,Y) \le 1$$

Example: 10.12 Let us calculate the correlation coefficient for the number of yellow and blue smarties in the box, of n smarties!
For that we need besides the covariance we have calculated, the variances.
To calculate variances, use the same trick of summing over indicator random variables.

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\sum_{i=1}^{n} Y_i\right)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(Y_i) \quad \text{(by independance of } Y_i\text{)}$$

$$= \sum_{i=1}^{n} p_y(1-p_y) \quad \text{by example 7.21}$$

$$= n\, p_y(1-p_y)$$

Similarly $Var(B) = n p_b (1-p_b)$

Putting these in the definition of correlation coefficient means

$$\rho(Y, B) = \frac{Cov[Y, B]}{\sqrt{Var(Y) Var(B)}}$$

$$= \frac{-n p_Y p_B}{\sqrt{n p_Y (1-p_Y) n p_b (1-p_b)}}$$

$$= \sqrt{\frac{p_Y p_b}{(1-p_Y)(1-p_b)}}$$

An extra property of covariance

$$
\begin{aligned}
\mathrm{Cov}(rX+s,\, tY+u) &= E\big[(rX+s-E[rX+s])(tY+u-E[tY+u])\big] \\
&= E\big[(rX+\cancel{s}-rE[X]-\cancel{s})(tY\cancel{+u}-tE[Y]-\cancel{u})\big] \\
&= E\big[r(X-E[X])\cdot t(Y-E[Y])\big] \\
&= rt\,E\big[(X-E[X])(Y-E[Y])\big] \\
&= rt\,\mathrm{Cov}(X,Y)
\end{aligned}
$$

$$\Rightarrow \boxed{\mathrm{Cov}(rX+s,\, tY+u) = rt\,\mathrm{Cov}(X,Y)}$$