Defn: Two jointly distributed r.v.s X and Y are called uncorrelated if $\mathrm{Cov}(X,Y)=0$.

Remark: The term "uncorrelated" arises from the simple fact that $\mathrm{Cov}(X,Y)=0 \Leftrightarrow \mathrm{Corr}(X,Y)=0$.

Properties of Covariance and Variance

(0) Suppose X is a r.v. with finite variance, then $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.

(1) Suppose X and Y are jointly distributed r.v.s with finite means and finite covariance. ~~then~~ Then

$$\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X).$$

(Symmetry of covariance)

Proof: Follows from the commutativity of product on $\mathbb{R}$.

(2) Suppose X_1, X_2, Y are jointly distributed r.v.s with finite means such that for $i=1,2$, X_i and Y have finite covariance. Then for any $\alpha_1, \alpha_2 \in \mathbb{R}$, the r.v.s Y and $\alpha_1 X_1 + \alpha_2 X_2$ have finite covariance, and

$$\mathrm{Cov}(\alpha_1 X_1 + \alpha_2 X_2, Y) = \alpha_1 \mathrm{Cov}(X_1,Y) + \alpha_2 \mathrm{Cov}(X_2,Y).$$

(Bilinearity of covariance)

Proof: Take $\alpha_1, \alpha_2 \in \mathbb{R}$.

Then $|(\alpha_1 X_1 + \alpha_2 X_2) Y| \leqslant |\alpha_1| |X_1 Y| + |\alpha_2| |X_2 Y|$

X_1, Y have finite means and finite covariance

$\Rightarrow$ $X_1 Y$ has finite mean

$\Rightarrow$ $|X_1 Y|$ has finite mean.

Similarly $|X_2 Y|$ has finite mean.

Therefore, $|(\alpha_1 X_1 + \alpha_2 X_2) Y|$ has finite mean

$\Rightarrow$ $(\alpha_1 X_1 + \alpha_2 X_2) Y$ has finite mean.

Also $\alpha_1 X_1 + \alpha_2 X_2$ and Y both have finite means. Therefore they have finite covariance.

Hence $\text{Cov}(\alpha_1 X_1 + \alpha_2 X_2, Y)$

$$= E[(\alpha_1 X_1 + \alpha_2 X_2) Y] - E(\alpha_1 X_1 + \alpha_2 X_2) E(Y)$$

$$= \alpha_1 E(X_1 Y) + \alpha_2 E(X_2 Y) - [\alpha_1 E(X_1) + \alpha_2 E(X_2)] E(Y)$$

$$= \alpha_1 [E(X_1 Y) - E(X_1) E(Y)] + \alpha_2 [E(X_2 Y) - E(X_2) E(Y)]$$

$$= \alpha_1 \operatorname{Cov}(X_1, Y) + \alpha_2 \operatorname{Cov}(X_2, Y),$$

which finishes the proof.

Remark: We have used the thm stated in Pg 183 a few times in the proof of (2).

(3) Suppose X, Y_1, Y_2 are jointly distributed r.v.s with finite means such that for $j = 1, 2$, X and Y_j have finite covariance. Then for any $\beta_1, \beta_2 \in \mathbb{R}$, the r.v.s X and $\beta_1 Y_1 + \beta_2 Y_2$ have finite covariance, and

$$\operatorname{Cov}(X, \beta_1 Y_1 + \beta_2 Y_2) = \beta_1 \operatorname{Cov}(X, Y_1) + \beta_2 \operatorname{Cov}(X, Y_2).$$

(Bilinearity of covariance)

Proof: $(1) + (2) \Rightarrow (3)$.

Remark: Properties (1), (2) (and hence (3)) together translates to the following algebraic statement: "covariance is a symmetric and bilinear form defined on an appropriate space".

(4) Suppose $X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_n$ are jointly distributed r.v.s with finite means such that X_i and Y_j have finite covariance for each pair $(i,j) \in \{1,2,\ldots,m\} \times \{1,2,\ldots,n\}$.

Then for all $\alpha_1, \alpha_2, \ldots, \alpha_m, \beta_1, \beta_2, \ldots, \beta_n \in \mathbb{R}$, the r.v.s $\sum_{i=1}^{m} \alpha_i X_i$ and $\sum_{j=1}^{n} \beta_j Y_j$ ~~are~~

have finite covariance and

$$\mathrm{Cov}\left(\sum_{i=1}^{m} \alpha_i X_i, \sum_{j=1}^{n} \beta_j Y_j\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i \beta_j \,\mathrm{Cov}(X_i, Y_j).$$

(<u>Bilinearity of covariance in its most general form</u>)

<u>**Proof (Sketch)**</u>: • First fix $m \in \mathbb{N}$, and show that

$$\mathrm{Cov}\left(\sum_{i=1}^{m} \alpha_i X_i, \sum_{j=1}^{n} \beta_j Y_j\right) = \sum_{j=1}^{n} \beta_j \,\mathrm{Cov}\left(\sum_{i=1}^{m} \alpha_i X_i, Y_j\right)$$

using induction on $n \in \mathbb{N}$ and (3).

• Then show that for each fixed $j \in \{1, 2, \ldots, n\}$,

$$\mathrm{Cov}\left(\sum_{i=1}^{m} \alpha_i X_i, Y_j\right) = \sum_{i=1}^{m} \alpha_i \,\mathrm{Cov}(X_i, Y_j)$$

using induction on $m \in \mathbb{N}$ and (2).

<u>Exc</u>: Convert the above sketch into a complete proof of (4).

Remarks ① The method of proof of (4) (as sketched in Pg 192) has a fancy name: two-fold ~~indue~~ induction on m and n.

② We shall see the importance of (4) as we write down its consequences and eventually use them in computation of variance and covariance.

(5) Under the assumptions of (4),

$$\mathrm{Cov}\left(\sum_{i=1}^{m} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathrm{Cov}(X_i, Y_j).$$

(Biadditivity of covariance)

Proof: Use (4) with $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 1$ and $\beta_1 = \beta_2 = \cdots = \beta_n = 1$.

(6) Suppose $X_1, X_2, \ldots, X_m$ are jointly distributed r.v.s with finite second moments ~~such~~ (so that ~~for all~~ each (i,j) with $1 \leq i < j \leq m$, X_i and X_j have finite covariance). Then for all $\alpha_1, \alpha_2, \ldots, \alpha_m \in \mathbb{R}$, the r.v. $\sum_{i=1}^{m} \alpha_i X_i$ has finite variance and

$$\mathrm{Var}\left(\sum_{i=1}^{m} \alpha_i X_i\right) = \sum_{i=1}^{m} \alpha_i^2 \, \mathrm{Var}(X_i) + 2 \sum_{\substack{i=1 \\ i<j}}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j \, \mathrm{Cov}(X_i, X_j).$$

Proof: Exc [Hint: $(0) + (4)_{\wedge}^{+(1)} \Rightarrow (6)$]

(7) Under the assumptions of (6),

$$\mathrm{Var}\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1}^{m} \mathrm{Var}(X_i) + 2\sum_{1 \le i < j \le m} \mathrm{Cov}(X_i, X_j).$$

Proof: Follows from (6) using $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 1$.

(8) Under the assumptions of (6),

$$(\text{av}) \ldots \quad \mathrm{Var}\left(\sum_{i=1}^{m} \alpha_i X_i\right) = \sum_{i=1}^{m} \alpha_i^2 \, \mathrm{Var}(X_i)$$

provided $X_1, X_2, \ldots, X_m$ are pairwise ~~uncorrelated~~ uncorrelated, i.e., for all (i,j) with $1 \le i < j \le m$, $\mathrm{Cov}(X_i, X_j) = 0$. In particular, if $X_1, X_2, \ldots, X_m$ are independent, then (av) holds.

Proof: $(6) \Rightarrow (8)$.

(9) For pairwise uncorrelated r.v.s $X_1, X_2, \ldots, X_m$ satisfying the assumptions of (6), the following special case of (av) holds:

$$(\text{avs}) \ldots \quad \mathrm{Var}\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1}^{m} \mathrm{Var}(X_i).$$

In particular, if $X_1, X_2, \ldots, X_m$ are ind, then (avs) holds.

Proof: Follows from (8) using $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 1$.

Exc: For top-to-random shuffling. (see Pg 163 – 168) compute $Var(T)$ using (9). In particular, recall the remark stated at the end of Pg 167.

Example: (Maxwell-Boltzmann Statistics)

Suppose $r \,(\in \mathbb{N})$ distinguishable particles are placed at random in $n \,(\in \mathbb{N})$ distinct (in terms of various physical characteristics) cells whose capacities are unlimited. Let N_0 be the ~~th~~ number of empty cells. Compute $E(N_0)$ and $Var(N_0)$. For the latter, assume $n \geq 2$.

Solution: Let us name the cells as Cell 1, Cell 2, ..., Cell n. Let, for each $j \in \{1, 2, \ldots, n\}$,

$$I_j = \begin{cases} 1 & \text{if Cell } j \text{ is empty,} \\ 0 & \text{if Cell } j \text{ is non-empty} \end{cases}$$

be the indicator that Cell j is empty.

Then clearly each $I_j \sim Ber(P(I_j = 1))$ even though $I_1, I_2, \ldots, I_n$ are not ind.

Therefore, for each $j \in \{1, 2, \ldots, n\}$,

$$E(I_j) = P(I_j = 1)$$

$$= P(\text{Cell } j \text{ is empty})$$

$$= \frac{(n-1)^r}{n^r}, \quad \text{and}$$

$$Var(I_j) = P(I_j = 1)(1 - P(I_j = 1))$$

$$= \frac{(n-1)^r}{n^r}\left(1 - \frac{(n-1)^r}{n^r}\right).$$

~~Therefore~~ Note that $N_0 = \sum_{j=1}^{n} I_j$.

Hence by linearity of expectation,

$$E(N_0) = \sum_{j=1}^{n} E(I_j) = n \cdot \frac{(n-1)^r}{n^r}$$

$$= \frac{(n-1)^r}{n^{r-1}}.$$

On the other hand, using (7) of Pg 194, we get

$$Var(N_0) = Var\left(\sum_{j=1}^{n} I_j\right)$$

$$\overset{(7)}{=} \sum_{j=1}^{n} Var(I_j) + 2 \sum_{1 \leq j < k \leq n} Cov(I_j, I_k)$$

$$= n \cdot \frac{(n-1)^r}{n^r}\left(1 - \frac{(n-1)^r}{n^r}\right) + 2 \sum_{1 \leq j < k \leq n} Cov(I_j, I_k).$$

Fix a pair (j, k) such that $1 \leq j < k \leq n$.

Then $Cov(I_j, I_k)$

$$= E(I_j I_k) - E(I_j)E(I_k)$$

$$= P(I_j = 1, I_k = 1) - \left[\frac{(n-1)^r}{n^r}\right]^2$$

[Since $I_j I_k \sim Ber(p)$, where $p = P(I_j = I_k = 1)$]

$$= P(\text{Cell } j \text{ and Cell } k \text{ are both empty}) - \left[\frac{(n-1)^r}{n^r}\right]^2$$

$$= \frac{(n-2)^r}{n^r} - \frac{(n-1)^{2r}}{n^{2r}}$$

$$= \frac{n^r(n-2)^r - (n-1)^{2r}}{n^{2r}} .$$

Exc: Show that for each pair (j,k) with $1 \leq j < k < n$, $\mathrm{Cov}(I_j, I_k) < 0$. Justify the sign of this covariance intuitively.

Going back to the calculation,

$\mathrm{Var}(N_0)$

$$= \frac{(n-1)^r}{n^{r-1}}\left(1 - \frac{(n-1)^r}{n^r}\right) + 2\sum_{1 \leq j < k \leq n} \mathrm{Cov}(I_j, I_k)$$

$$= \frac{(n-1)^r}{n^{r-1}}\left(1 - \frac{(n-1)^r}{n^r}\right) + 2\,\frac{n(n-1)}{2} \cdot \frac{n^r(n-2)^r - (n-1)^{2r}}{n^{2r}}$$

$$= \frac{(n-1)^r}{n^{r-1}}\left(1 - \frac{(n-1)^r}{n^r}\right) + \frac{n^r(n-1)(n-2)^r - (n-1)^{2r+1}}{n^{2r-1}}$$

Exc: Consider a Maxwell-Boltzmann model with $r \geq 1$ particles and $n \geq 3$ cells. Let N_0 be the number of empty cells and N_1 be the number of cells that contain exactly one particle. Show that

$$\mathrm{Cov}(N_0, N_1) = \frac{r(n-1)(n-2)^{r-1}}{n^{r-1}} - \frac{r(n-1)^{2r-1}}{n^{2r-2}}.$$

[**Hint:** Use (5) of Pg 193.]

Exc: Suppose there are N envelopes with N distinct addresses written on them. There are N letters that ~~has~~ need to be sent to these N addresses in a coherent manner. Suppose a sleepy secretary ~~randomly~~ puts the N letters into the N envelopes at random (so that each envelope contains exactly one letter). Let X be the number of letters that are put inside the correct envelope. Compute $E(X)$ and $\mathrm{Var}(X)$.