Lecture 18

Addition rule

Revision

Covariance

Correlation

Examples

Addition rule for variances:

$X_1, X_2, \ldots, X_n$ independent implies

$$Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} Var(X_k)$$

- Sufficient to prove for $n = 2$. Set $X = X_1,\ Y = X_2$. Then

$$\begin{aligned}
Var(X+Y) &= E\left\{[(X+Y) - E\{X+Y\}]^2\right\} \\
&= E\left\{[(X+Y) - (\mu_X + \mu_Y)]^2\right\} \\
&= E\left\{[(X-\mu_X) - (Y-\mu_Y)]^2\right\} \\
&= E\left\{(X-\mu_X)^2 + (Y-\mu_Y)^2 - 2(X-\mu_X)(Y-\mu_Y)\right\} \\
&= Var(X) + Var(Y) - 2E\{X-\mu_X\} \times E\{Y-\mu_Y\} \\
&= Var(X) + Var(Y)
\end{aligned}$$

$$\boxed{X \sim \text{Bin}(n,p) \;\Rightarrow\; Var(X) = npq,\; SD(X) = \sqrt{npq}.}$$

- $X \sim \text{Bin}(n,p) \;\Rightarrow\; X = I_1 + I_2 + \ldots\; I_n$ where the I_k are <u>independent</u> indicator variables and $P\{I_k = 1\} = p$
- Recall $E\{I_k\} = p,\; Var(I_k) = pq.$
- Therefore, by the addition rule,

$$Var(X) \;=\; Var(I_1 + I_2 + \ldots\; I_n) \;=\; npq$$

which is a lot easier than

$$\sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} \;-\; (np)^2$$

and even easier than

$$npq + n^2p^2 - (np)^2$$

Revision

- $Var(X) \equiv E\{(X-\mu)^2\} \equiv E\{X^2\} \; - \; [E\{X\}]^2$
- $SD(X) \equiv \sqrt{Var(X)}$
- Indicators: $Var(I_A) \; = \; pq$
- $Var(aX+b) \; = \; a^2\, Var(X)$
- $SD(aX+b) \; = \; a\, SD(X)$
- Standardisation: $X^* \equiv (X-\mu)/\sigma$
  $\Rightarrow \qquad E\{X^*\} = 0, \; Var(X^*) = 1$
- Chebychev's Inequality: $P\{|X-\mu| \geq k\,\sigma\} \; \leq \; 1/k^2$
- Addition rule for variances: $X_1, X_2, \ldots, X_n$ independent
  $\Rightarrow \qquad Var\left(\sum_{k=1}^n X_k\right) \; = \; \sum_{k=1}^n Var(X_k)$

Covariance

$$Cov(X,Y) \;=\; E\{(X-\mu_X)\,(Y-\mu_Y)\}$$

where $\mu_X = E\{X\}, \quad \mu_Y = E\{Y\}$.

Alternative formula:

$$\begin{aligned} Cov(X,Y) &= E\{(X-\mu_X)\,(Y-\mu_Y)\} \\ &= E\{XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y\} = E\{XY\} - \mu_X\mu_Y \end{aligned}$$

Variance of sum:

$$\begin{aligned} Var(X+Y) &= E\{[X+Y \;-\; (\mu_X+\mu_Y)]^2\} \\ &= E\{[X-\mu_X]^2 + [Y-\mu_Y]^2 + 2(X-\mu_X)\,(Y-\mu_Y)\} \\ &= Var(X) \;+\; Var(Y) \;+\; 2Cov(X,Y) \end{aligned}$$

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

$$\begin{aligned} Cov(X,Y) &= E\{(X-\mu_X)(Y-\mu_Y)\} \\ &= E\{X-\mu_X\} \times E\{Y-\mu_Y\} \;=\; 0 \times 0 \end{aligned}$$

BUT: $Cov(X, Y) = 0 \not\Rightarrow X, Y$ independent!

Y\ X	-1	0	$+1$	
-1	$--$	$1/4$	$--$	$1/4$
0	$1/4$	$--$	$1/4$	$1/2$
$+1$	$--$	$1/4$	$--$	$1/4$
	$1/4$	$1/2$	$1/4$	1

$Cov(X, Y) = 0$, but X and Y are <u>obviously</u> dependent.

Covariance of indicators

$$
\begin{aligned}
I_A I_B &= I_{AB} = I_{A\cap B} \\
\Rightarrow \qquad Cov(I_A, I_B) &= E\{I_A I_B\} - E\{I_A\}E\{I_B\} \\
&= P\{AB\} - P\{A\}P\{B\}
\end{aligned}
$$

The relationship between A and B is that of

positive dependence	$\Longleftrightarrow P\{AB\} > P\{A\}P\{B\}$ $\Longleftrightarrow P\{A\|B\} > P\{A\}$
independence	$\Longleftrightarrow P\{AB\} = P\{A\}P\{B\}$ $\Longleftrightarrow P\{A\|B\} = P\{A\}$
negative dependence	$\Longleftrightarrow P\{AB\} < P\{A\}P\{B\}$ $\Longleftrightarrow P\{A\|B\} < P\{A\}$

Correlation

$$Corr(X,Y) = \frac{Cov(X,Y)}{SD(X)\,SD(Y)}$$

Correlations do not feel standardisation:

$$\begin{aligned}
Corr(X^*,Y^*) &= \frac{Cov(X^*,Y^*)}{SD(X^*)\,SD(Y^*)} \\
&= \frac{Cov\left(\frac{X-\mu_X}{SD(X)}, \frac{Y-\mu_Y}{SD(Y)}\right)}{SD(X^*)\,SD(Y^*)} \\
&= \frac{E\left\{\frac{(X-\mu_X)(Y-\mu_Y)}{SD(X)SD(Y)}\right\}}{1} \\
&= \frac{E\{(X-\mu_X)(Y-\mu_Y)\}}{SD(X)SD(Y)} \\
&= Corr(X,Y)
\end{aligned}$$

Correlations are <u>always</u> between -1 and +1

PROOF: If X and Y are any two random variables, it suffices to show this holds for X^* and Y^*.

$$\begin{aligned} 0 &\leq E\{(X^* - Y^*)^2\} = 1 + 1 - 2E\{X^*Y^*\} \Rightarrow E\{X^*Y^*\} \leq 1 \\ 0 &\leq E\{(X^* + Y^*)^2\} = 1 + 1 + 2E\{X^*Y^*\} \Rightarrow E\{X^*Y^*\} \geq -1 \end{aligned}$$

Correlations of ± 1

$$\begin{aligned} Corr(X, Y) = +1 \quad &\Longleftrightarrow \quad Corr(X^*, Y^*) = +1 \\ &\Longleftrightarrow \quad E\{(X^* - Y^*)^2\} = 0 \\ &\Longleftrightarrow \quad X^* = Y^* \\ &\Longleftrightarrow \quad X = aY + b \end{aligned}$$

for some $a > 0$.

Correlations are (almost) unaffected by linear transformations

- PROOF 1: Go through the same proof used to show $Corr(X, Y) = Corr(X^*, Y^*)$.

- PROOF 2: Be clever, and note:

$$\begin{aligned} Corr(aX + b, cY + d) &= Corr((aX + b)^*, (cY + d)^*) \\ &= \text{sign}(ac) \times Corr(X^*, Y^*) \\ &= \text{sign}(ac) \times Corr(X, Y) \end{aligned}$$

Example: Betting on $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$

- Gambler chooses a set $A \subset \Omega$, containing <u>four</u> numbers. (e.g. $A = (1, 2, 3, 4)$.)
- She places a \$1 "even money" bet on a A. i.e. She wins \$1 if A occurs. Her <u>net</u> gain is $X = 2I_A - 1$.
- Since $E\{X\} = 0$ the game is "fair".
- She now chooses another set $B \subset \Omega$, also containing four numbers, and bets on these as well. Set $Y = 2I_B - 1$.
- Total gain is $S = X + Y$, which still has $E\{S\} = 0$.

$$\begin{aligned} Var(S) &= Var(X) + Var(Y) + 2\,Cov(X, Y) \\ &= 2 + 2\,Corr(X, Y) \\ &= 2 + 2\,Corr(I_A, I_B) \end{aligned}$$

$$\boxed{Var(S) \;=\; 2 + 2\,Corr(I_A, I_B)}$$

$$\begin{aligned} Corr(I_A, I_B) &= \frac{E\{I_A I_B\} - E\{I_A\}E\{I_B\}}{SD(I_A)SD(I_B)} \\ &= \frac{P\{AB\} - 1/4}{1/4} \;=\; -1 + \#(AB)/2 \\ \Rightarrow \qquad & Var(S) \;=\; \#(AB) \end{aligned}$$

$\#(AB)$	$Corr(X,Y)$	$Var(S)$	Comment
0	-1	0	No risk: Loss = Win
1	$-1/2$	1	
2	0	2	X and Y are independent
3	$1/2$	3	
4	1	4	Highest risk