DIS 11A

COVARIANCE

what is the association between two RVs?

$$\operatorname{cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \qquad \mu_X = E[X]$$

$$= E[XY] - E[X]E[Y] \qquad \mu_Y = E[Y]$$

some useful properties:

1. X, Y independent $\Rightarrow \operatorname{cov}(X,Y) = 0$ (converse is NOT true)
2. $\operatorname{cov}(X,X) = \operatorname{var}(X)$
3. covariance is bilinear

   ex. $\operatorname{Cov}(a_1X_1 + a_2X_2,\ b_1Y_1 + b_2Y_2)$

   $= a_1b_1\operatorname{Cov}(X_1,Y_1) + a_2b_1\operatorname{Cov}(X_2,Y_1)$

   $+ a_1b_2\operatorname{Cov}(X_1,Y_2) + a_2b_2\operatorname{Cov}(X_2,Y_2)$

4. $\operatorname{var}(X+Y) = \operatorname{var}(X) + \operatorname{var}(Y) + 2\operatorname{cov}(X,Y)$

CORRELATION

a more interpretable measure of association

$$\text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\sigma(X)\,\sigma(Y)}$$

where $\sigma(X) = \sqrt{\text{var}(X)}$

why is this more interpretable?

↳ while the range of values of covariance varies between variables, correlation is always between $\underline{-1}$ and $\underline{+1}$

CONDITIONAL EXPECTATION

$$E[X \mid Y=y] = \sum_{x} x \times P(X=x \mid Y=y)$$

LAW OF ITERATED EXPECTATION

$$E[X] = E[E[X \mid Y]] = \sum_{y} E[X \mid Y=y]\, P(Y=y)$$

(proof)

$$= \sum_{y} \sum_{x} x \times P(X=x \mid Y=y)\, P(Y=y)$$

$$= \sum_{y} \sum_{x} x \times P(X=x, Y=y)$$

$$= \sum_{x} x \sum_{y} P(X=x, Y=y)$$

$$= \sum_{x} x\, P(X=x)$$

INDICATOR VARIABLES

$$I_A = \begin{cases} 1 & \text{if } A \text{ happens} \\ 0 & \text{otherwise} \end{cases}$$

$$E[I_A] = P(A) \xrightarrow{I_A^2 = I_A} E[I_A^2] = P(A)$$

$$E[(I_1 + \dots + I_n)^2] = E[(I_1 + \dots + I_n)(I_1 + \dots + I_n)]$$

$$= E\left[\sum_{i=1}^{n} I_i^2 + \sum_{i \neq j}^{n} I_i I_j\right]$$

$$= \sum_{i=1}^{n} E[I_i^2] + \sum_{i \neq j}^{n} E[I_i I_j]$$