Expectation from joint pdfs/pmfs

- The expected values of **X** and **Y** are
  $E[\mathbf{X}] = \int_{-\infty}^{\infty} u \cdot f_{\mathbf{X}}(u)\,du$; $E[\mathbf{Y}] = \int_{-\infty}^{\infty} v \cdot f_{\mathbf{Y}}(v)\,dv$
  for continuous random variables and by
  $E[\mathbf{X}] = \sum u_i \cdot p_{\mathbf{X}}(u_i)$; $E[\mathbf{Y}] = \sum v_i \cdot p_{\mathbf{Y}}(v_i)$ for discrete random variables
- Given the joint pdf or pmf of **X** and **Y**, we can first compute the (marginal) pdf or pmf of **X** or **Y** and substitute in the above

Doing everything in one step

- $E[\mathbf{X}] = \int_{-\infty}^{\infty} u \cdot f_{\mathbf{X}}(u)\,du$; $E[\mathbf{Y}] = \int_{-\infty}^{\infty} v \cdot f_{\mathbf{Y}}(v)\,dv$
- $f_{\mathbf{X}}(u) = \int_{v=-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)\,dv$; $f_{\mathbf{Y}}(v) = \int_{u=-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)\,du$
- $E[\mathbf{X}] = \int_{-\infty}^{\infty} u \cdot \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)\,dv\,du$
  $= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u \cdot f_{\mathbf{X},\mathbf{Y}}(u,v)\,dv\,du$
  $= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u \cdot f_{\mathbf{X},\mathbf{Y}}(u,v)\,du\,dv$

It works the same way for Y too!

- $E[\mathbf{X}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u \cdot f_{\mathbf{X},\mathbf{Y}}(u,v)\,du\,dv$
- Similarly,
  $E[\mathbf{Y}] = \int_{-\infty}^{\infty} v \cdot \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)\,du\,dv$
  $= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v \cdot f_{\mathbf{X},\mathbf{Y}}(u,v)\,du\,dv$
  $= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v \cdot f_{\mathbf{X},\mathbf{Y}}(u,v)\,dv\,du$
- For discrete RVs, integrals become sums

Generalization to n variables

- Given the joint pdf $f_{\underline{\mathbf{X}}}(\underline{u})$ or pmf $p_{\underline{\mathbf{X}}}(\underline{u})$ of n random variables $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n$, $E[\mathbf{X}_i]$ is given by the n-dimensional integral of $u_i \cdot f_{\underline{\mathbf{X}}}(\underline{u})$ over the entire space or the n-fold sum of $u_i \cdot p_{\underline{\mathbf{X}}}(\underline{u})$
- Expectation of a vector
- If $\underline{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n)$, then $E[\underline{\mathbf{X}}] = (E[\mathbf{X}_1], E[\mathbf{X}_2], \ldots E[\mathbf{X}_n])$
- The expectation of a random vector is the vector of the expectations

LOTUS works in the same way…

- Joint pdf $f_{\underline{\mathbf{X}}}(\underline{u})$ or pmf $p_{\underline{\mathbf{X}}}(\underline{u})$ of n random variables $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n$
- $E[g(\mathbf{X}_i)]$ is given by the (n-dimensional) integral of $g(u_i) \cdot f_{\underline{\mathbf{X}}}(\underline{u})$ over the entire space or the n-fold sum of $g(u_i) \cdot p_{\underline{\mathbf{X}}}(\underline{u})$
- This is just LOTUS with the calculation of the marginal pdf of $\mathbf{X}_i$ being merged with the calculation of $E[g(\mathbf{X}_i)]$ into one giant step for mankind

Generalization of LOTUS

- Joint pdf $f_{\underline{\mathbf{X}}}(\underline{u})$ or pmf $p_{\underline{\mathbf{X}}}(\underline{u})$
- $g(\underline{\mathbf{X}}) = g(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ is a function of the n random variables
- Generalized LOTUS or gLOTUS
- $E[g(\underline{\mathbf{X}})] = E[g(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_1)]$ can be computed as the n-dimensional integral of $g(\underline{u}) \cdot f_{\underline{\mathbf{X}}}(\underline{u})$ or the n-fold sum of $g(\underline{u}) \cdot p_{\underline{\mathbf{X}}}(\underline{u})$
- gLOTUS is just like LOTUS: Multiply the function by the pdf (or pmf) and integrate (or sum) over all the variables

Expectation of a sum

- If $g(\underline{X}) = \mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n$, then $E[g(\underline{X})] = E[\mathbf{X}_1 + \mathbf{X}_2 + \ldots + \mathbf{X}_n]$

$= \int\int\ldots\int (u_1+u_2+\ldots+u_n)\bullet f_{\underline{X}}(\underline{u})$

$= \int\int\ldots\int u_1\bullet f_{\underline{X}}(\underline{u}) + u_2\bullet f_{\underline{X}}(\underline{u}) + \ldots + u_n\bullet f_{\underline{X}}(\underline{u})$

$= \int\int\ldots\int u_1\bullet f_{\underline{X}}(\underline{u}) + \ldots + \int\int\ldots\int u_n\bullet f_{\underline{X}}(\underline{u})$

$= E[\mathbf{X}_1] + E[\mathbf{X}_2] + \ldots + E[\mathbf{X}_n]$

Saying it in words …

- $E[\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n] = E[\mathbf{X}_1]+E[\mathbf{X}_2]+\ldots+E[\mathbf{X}_n]$
- Expectation of sum = sum of expectations
- This applies to all random variables
- RVs need not be independent or Gaussian or all discrete or all continuous or …
- More generally, $E\left[\sum a_i\bullet\mathbf{X}_i\right] = \sum a_i\bullet E[\mathbf{X}_i]$
- $\sum a_i\bullet\mathbf{X}_i = \underline{\mathbf{X}}\bullet\underline{a}^T = [\mathbf{X}_1,\ldots,\mathbf{X}_n]\bullet[a_1,\ldots,a_n]^T$
- $E[\underline{\mathbf{X}}\bullet\underline{a}^T] = E[\underline{\mathbf{X}}]\bullet\underline{a}^T = \underline{\mu}_{\mathbf{X}}\bullet\underline{a}^T$
- Expectation is a linear operator

Does it work for all functions? No!

- $E[\mathbf{X}_1+\mathbf{X}_2+\ldots+\mathbf{X}_n] = E[\mathbf{X}_1]+E[\mathbf{X}_2]+\ldots+ E[\mathbf{X}_n]$
- Don't extend this to other functions, e.g., $E[\mathbf{X}_1\mathbf{X}_2\ldots\mathbf{X}_n] \neq E[\mathbf{X}_1]E[\mathbf{X}_2]\ldots E[\mathbf{X}_n]$
- But, if the $\mathbf{X}_1$'s are independent RVs, then

$E[\mathbf{X}_1\mathbf{X}_2\ldots\mathbf{X}_n] = \int\int\ldots\int u_1u_2\ldots u_n\bullet f_{\underline{X}}(\underline{u})$

$= \int u_1 f_{X_1}(u_1) \int u_2 f_{X_2}(u_2) \ldots \int u_n f_{X_n}(u_n)$

$= E[\mathbf{X}_1]\,E[\mathbf{X}_2]\ldots E[\mathbf{X}_n]$

Independence makes life so easy…

- $E[\mathbf{X}_1\mathbf{X}_2\ldots\mathbf{X}_n] = E[\mathbf{X}_1]E[\mathbf{X}_2]\ldots E[\mathbf{X}_n]$ if the $\mathbf{X}_1$'s are independent RVs
- More generally, if the $\mathbf{X}_1$'s are independent RVs, then

  $E[g_1(\mathbf{X}_1)g_2(\mathbf{X}_2)\ldots g_n(\mathbf{X}_n)]$
  $= E[g_1(\mathbf{X}_1)]E[g_2(\mathbf{X}_2)] \ldots E[g_n(\mathbf{X}_n)]$
- More strongly, $g_1(\mathbf{X}_1), g_2(\mathbf{X}_2), \ldots, g_n(\mathbf{X}_n)$ also are independent random variables
- Why is it so? Why are the $g_i(\mathbf{X}_i)$ independent random variables?

Proof by contradiction

- Prove: If **X** and **Y** are independent, then so are g(**X**) and h(**Y**)
- Independence of **X** and **Y**: knowing the value taken on by **X** tells us nothing new (nothing that we did not already know) about the value taken on by **Y**
- If knowing the value taken on by g(**X**) told us something about the value taken on by h(**Y**), then knowledge of **X** could allow us to infer something about **Y**! Contradiction!

Covariance of X and Y

- For independent RVs **X** and **Y**,

  $E[\mathbf{XY}] = E[\mathbf{X}]E[\mathbf{Y}]$
- But, in general, $E[\mathbf{XY}] \neq E[\mathbf{X}]E[\mathbf{Y}]$
- The difference $E[\mathbf{XY}] - E[\mathbf{X}]E[\mathbf{Y}]$ is called the covariance of **X** and **Y**
- $\text{cov}(\mathbf{X}, \mathbf{Y}) = E[\mathbf{XY}] - E[\mathbf{X}]E[\mathbf{Y}]$
- The following is a more common definition: $\text{cov}(\mathbf{X}, \mathbf{Y}) = E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{Y} - E[\mathbf{Y}])]$
- Exercise: use linearity of the expectation operator to show equality of the formulas

Some thoughts about covariance

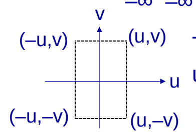
- Covariance is a pair-wise property: we do not talk of covariance of 3 (or more) RVs
- cov(**X**, **Y**) = E[**XY**] – E[**X**]E[**Y**] = cov(**Y**, **X**)
- Covariance generalizes the notion of the variance of a random variable: covariance of **X** with itself is the variance of **X**
- cov(**X**, **X**) = E[(**X** – E[**X**])(**X** – E[**X**])] = E[(**X** – E[**X**])^2] = $(\sigma_X)^2$ = var(**X**)
- Independent random variables have zero covariance

The correlation coefficient

- var(**X**) = $(\sigma_X)^2$; var(**Y**) = $(\sigma_Y)^2$ are finite
  $-\sigma_X \bullet \sigma_Y \le$ cov(**X**, **Y**) $\le +\sigma_X \bullet \sigma_Y$
- Probabilistic version of Schwarz Inequality
- Ratio cov(**X**, **Y**)/$(\sigma_X \bullet \sigma_Y)$ is the correlation coefficient of **X** and **Y**, and is denoted by $\rho_{X,Y}$ or simply ρ if no confusion can arise
- If $\rho_{X,Y} = \pm 1$, **X** and **Y** are called perfectly (positively or negatively) correlated RVs
- If $\rho_{X,Y} = 0$, **X** and **Y** are called uncorrelated random variables

What does covariance measure?

- Assume E[**X**] = E[**Y**] = 0 for simplicity so that cov(**X**, **Y**) = E[**XY**]
- E[**XY**] = $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u \bullet v \bullet f_{X,Y}(u,v)\,dv\,du$



The corners contributes uv•$[f_{X,Y}(u,v) + f_{X,Y}(-u,-v) - f_{X,Y}(u,-v) - f_{X,Y}(-u,v)]$• dv•du to the integral

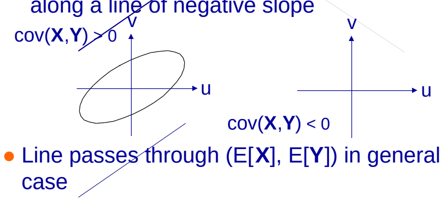
- E[**XY**] = Integral over first quadrant only

Where is the mass spread?

- cov(**X**, **Y**) = E[**XY**] is the integral of uv•$[f_{X,Y}(u,v)+f_{X,Y}(-u,-v)-f_{X,Y}(u,-v)-f_{X,Y}(-u,v)]$ over the first quadrant
- We are taking the difference between the masses in 1st plus 3rd quadrant and the 2nd plus 4th quadrants (with weighting factor uv)
- Roughly speaking, cov(**X**, **Y**) > 0 means more mass in 1st plus 3rd quadrants than in 2nd plus 4th quadrants

Positive or negative slope?

- If cov(**X**, **Y**) > 0, the probability masses lie along a line of positive slope
- If cov(**X**, **Y**) < 0, the probability masses lie along a line of negative slope



- Line passes through (E[**X**], E[**Y**]) in general case

The variance of X ± Y

- E[**X**] = μ_X; E[**Y**] = μ_Y; E[**X** ± **Y**] = $\mu_X \pm \mu_Y$
- var(**X** ± **Y**) = E[(**X** ± **Y** – $\{\mu_X \pm \mu_Y\}$)^2]
  = E[($\{$**X**– $\mu_X\}$ ± $\{$**Y** –$\mu_Y\}$)^2]
  = E[(**X**– μ_X)^2 + (**Y** –μ_Y)^2 ± 2(**X**– μ_X)(**Y** –μ_Y)]
  = var(**X**) + var(**Y**) ± 2•cov(**X**,**Y**)
  = $(\sigma_X)^2 + (\sigma_Y)^2 \pm 2\bullet\rho\bullet\sigma_X\bullet\sigma_Y$
- More generally, var(a**X** + b**Y**)
  = $a^2\bullet(\sigma_X)^2 + b^2\bullet(\sigma_Y)^2$ + 2•a•b•cov(**X**,**Y**)
  = $a^2\bullet(\sigma_X)^2 + b^2\bullet(\sigma_Y)^2 \pm 2\bullet a\bullet b\bullet\rho\bullet\sigma_X\bullet\sigma_Y$

The Cauchy-Schwarz Inequality

- The variance of a random variable cannot be negative
- var(a**X** ± b**Y**) $= a^2 \bullet (\sigma_X)^2 + b^2 \bullet (\sigma_Y)^2 \pm 2 \bullet a \bullet b \bullet \rho \bullet \sigma_X \bullet \sigma_Y$
- Now, choose $a = 1/\sigma_X$ and $b = 1/\sigma_Y$
- var($\mathbf{X}/\sigma_X \pm \mathbf{Y}/\sigma_Y$) = 1 + 1 ± 2•ρ ≥ 0
- 2(1 ± ρ) ≥ 0   1 + ρ ≥ 0 and 1 – ρ ≥ 0
- Conclusion: –1 ≤ ρ ≤ 1
- If ρ = ±1, all the probability mass lies on a straight line in the plane and **Y** = α+β**X**

Special values of ρ

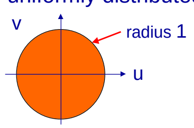
- If ρ = ±1, all the probability mass lies on a straight line in the plane and **Y** = α+β**X**
- If ρ = 0, **X** and **Y** are said to be uncorrelated random variables
- Independent random variables have zero covariance and hence are uncorrelated
- But, uncorrelated random variables are not necessarily independent
- Independence is a very strong property

Independent vs uncorrelated – I

- If **X** and **Y** are independent RVs, then they are also uncorrelated random variables
- Independence is a very strong property: the joint pdf/pmf factors into the product of the marginal pdfs/pmfs at every point in the plane
- ρ = 0 merely means that an expectation integral or sum is 0
- Average = 0 can hold even if the random variables are dependent

Independent vs uncorrelated – II

- Let (**X**,**Y**) be uniformly distributed on the unit disc



- –1 ≤ **X**,**Y** ≤ +1, but knowing the value of **X** restricts **Y** to be in a smaller range
- **X** and **Y** are not independent RVs
- But, E[**X**] = E[**Y**] = E[**XY**] = 0 and hence **X** and **Y** are uncorrelated RVs

Independent vs uncorrelated – III

- Suppose **X** and **Y** are uncorrelated
- Then, var(**X** ± **Y**) = var(**X**) + var(**Y**)
- Note that variances always add even if the random variables are being subtracted
- This holds for independent RVs too
- The mean of a sum is the sum of the means holds for all random variables
- The variance of a sum is the sum of the variances holds for uncorrelated RVs (and for the subclass of independent RVs too)

Covariance is a bilinear function

- Bilinear means linear in both arguments
- cov(a**X**+b**Y**, **Z**) = a•cov(**X**,**Z**) + b•cov(**Y**,**Z**)
- cov(**X**, a**Y**+b**Z**) = a•cov(**X**,**Y**) + b•cov(**X**,**Z**)
- cov(a**X** + b**Y**, c**X** + d**Y**)
  = ac•cov(**X**, **X**) + bd•cov(**Y**, **Y**)
  + ad•cov(**X**, **Y**) + bc•cov(**Y**, **X**)
  = ac•var(**X**)+bd•var(**Y**)+(ad+bc)•cov(**X**, **Y**)
  $= ac \bullet (\sigma_X)^2 + bd \bullet (\sigma_Y)^2 + (ad+bc) \bullet \mathrm{cov}(\mathbf{X}, \mathbf{Y})$

Even more generally…

- $\mathrm{cov}(a\mathbf{X}+b\mathbf{Y}+e,\ c\mathbf{Z}+d\mathbf{W}+f)$
  $= ac\bullet\mathrm{cov}(\mathbf{X}, \mathbf{Z}) + bd\bullet\mathrm{cov}(\mathbf{Y}, \mathbf{W})$
  $+ ad\bullet\mathrm{cov}(\mathbf{X}, \mathbf{W}) + bc\bullet\mathrm{cov}(\mathbf{Y}, \mathbf{Z})$
- Note that $\mathrm{cov}(\mathbf{X}, \alpha) = 0$ for any random variable $\mathbf{X}$ and any constant α
- α does not vary and hence cannot co-vary with $\mathbf{X}$
- We now consider the case of n random variables

n random variables

- Given n random variables $\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n$ and $\underline{\mathbf{X}} = [\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n]$,
  $E[\underline{\mathbf{X}}] = [\ E[\mathbf{X}_1], E[\mathbf{X}_2], \ldots E[\mathbf{X}_n]\ ] = \underline{\mu}_{\mathbf{X}}$
- The expectation of a random vector is the vector of the expectations
- Expectation of a matrix with RVs as entries = matrix of expectations
- $\sum a_i\bullet\mathbf{X}_i = \underline{\mathbf{X}}\bullet\underline{a}^T = [\mathbf{X}_1,\ldots,\mathbf{X}_n]\bullet[a_1,\ldots,a_n]^T$
- $E[\underline{\mathbf{X}}\bullet\underline{a}^T] = E[\underline{\mathbf{X}}]\bullet\underline{a}^T = \underline{\mu}_{\mathbf{X}}\bullet\underline{a}^T$

The covariance matrix — I

- There are n^2 pairs of random variables $\mathbf{X}_i$ and $\mathbf{X}_j$ giving n^2 covariance functions
- n of these are $\mathrm{cov}(\mathbf{X}_i, \mathbf{X}_i) = \mathrm{var}(\mathbf{X}_i)$
- The covariance matrix R is a symmetric n×n matrix with i-jth entry $r_{i,j} = \mathrm{cov}(\mathbf{X}_i, \mathbf{X}_j)$
- The variances of the $\mathbf{X}_i$'s appear along the diagonal of the matrix
- Uncorrelated RVs R is diagonal matrix

The covariance matrix — II

- $\underline{\mathbf{X}} = [\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n]$ and $\underline{\mu}$ are 1×n matrices or row vectors
- $(\underline{\mathbf{X}} - \underline{\mu}_{\mathbf{X}})^T$ is a n×1 matrix or column vector
- $(\underline{\mathbf{X}} - \underline{\mu}_{\mathbf{X}})^T\bullet(\underline{\mathbf{X}} - \underline{\mu}_{\mathbf{X}})$ is a n×n matrix whose i-jth entry is $(\mathbf{X}_i-\mu_i)(\mathbf{X}_j-\mu_j)$
- E[matrix] = matrix of expectations
- $R = E[(\underline{\mathbf{X}} - \underline{\mu}_{\mathbf{X}})^T\bullet(\underline{\mathbf{X}} - \underline{\mu}_{\mathbf{X}})]$ is an n×n matrix with i-jth entry $r_{i,j} = \mathrm{cov}(\mathbf{X}_i, \mathbf{X}_j)$

The covariance matrix — III

- $R = E[(\underline{\mathbf{X}} - \underline{\mu})^T\bullet(\underline{\mathbf{X}} - \underline{\mu})]$ is an n×n matrix with i-jth entry $r_{i,j} = \mathrm{cov}(\mathbf{X}_i, \mathbf{X}_j)$
- R is a symmetric positive semidefinite (also known as symmetric nonnegative definite) matrix
- R is used to find variance and covariances of linear combinations of the $\mathbf{X}_i$
- Example:What is $\mathrm{var}\left(\sum a_i\bullet\mathbf{X}_i\right) = \mathrm{var}(\underline{\mathbf{X}}\bullet\underline{a}^T)$?

The variance of a sum

- $E\left[\sum a_i\bullet\mathbf{X}_i\right] = E[\underline{\mathbf{X}}\bullet\underline{a}^T] = E[\underline{\mathbf{X}}]\bullet\underline{a}^T = \underline{\mu}_{\mathbf{X}}\bullet\underline{a}^T$
- $\mathrm{var}\left(\sum a_i\bullet\mathbf{X}_i\right) = E\left[\left\{\sum (a_i\bullet\mathbf{X}_i - a_i\bullet\mu_i)\right\}^2\right]$

  $= E\left[\sum (a_i)^2(\mathbf{X}_i-\mu_i)^2 + \sum\sum a_i a_j(\mathbf{X}_i-\mu_i)\bullet(\mathbf{X}_j-\mu_j)\right]$

  $= \sum (a_i)^2\mathrm{var}(\mathbf{X}_i) + \sum\sum a_i\bullet a_j\bullet\mathrm{cov}(\mathbf{X}_i, \mathbf{X}_j)$

  where the second sum is over all $i \neq j$
- In terms of the covariance matrix R, $\mathrm{var}(\underline{a}\bullet\underline{\mathbf{X}}) = \underline{a}\bullet R\bullet\underline{a}^T$

Doing it entirely with matrices

- $\underline{\mathbf{X}}\bullet\underline{a}^T$ is a random variable with mean $\underline{\mu}_X\bullet\underline{a}^T$
- $\text{var}(\underline{\mathbf{X}}\bullet\underline{a}^T)$

  $= E[(\underline{\mathbf{X}}\bullet\underline{a}^T - \underline{\mu}_X\bullet\underline{a}^T)^2] = E[\{(\underline{\mathbf{X}} - \underline{\mu}_X)\bullet\underline{a}^T\}^2]$

  $= E[\{(\underline{\mathbf{X}} - \underline{\mu}_X)\bullet\underline{a}^T\}\bullet\{(\underline{\mathbf{X}} - \underline{\mu}_X)\bullet\underline{a}^T\}]$

  $= E[(\underline{a}^T)^T\bullet(\underline{\mathbf{X}} - \underline{\mu}_X)^T\bullet(\underline{\mathbf{X}} - \underline{\mu}_X)\bullet\underline{a}^T]$

  $= \underline{a}\bullet E[(\underline{\mathbf{X}} - \underline{\mu}_X)^T\bullet(\underline{\mathbf{X}} - \underline{\mu}_X)]\bullet\underline{a}^T$

  $= \underline{a}\bullet R\bullet\underline{a}^T$
- $\text{var}(\underline{\mathbf{X}}\bullet\underline{a}^T) = \underline{a}\bullet R\bullet\underline{a}^T$

Quadratic forms

- $\text{var}(\underline{\mathbf{X}}\bullet\underline{a}^T) = \underline{a}\bullet R\bullet\underline{a}^T$
- $\underline{a}\bullet R\bullet\underline{a}^T$ is called a quadratic form in the n variables $a_1,\ldots,a_n$ because the squares (but no higher powers) of the a_i occur in the expansion of $\underline{a}\bullet R\bullet\underline{a}^T$
- $\text{var}(\underline{\mathbf{X}}\bullet\underline{a}^T) \geq 0$ for all choices of $a_1,\ldots,a_n$ such that at least one a_i is nonzero
- Hence the name nonnegative definite

Covariance of linear combinations

- More generally, $\text{cov}(\underline{\mathbf{X}}\bullet\underline{a}^T, \underline{\mathbf{X}}\bullet\underline{b}^T) = \underline{a}\bullet R\bullet\underline{b}^T = \underline{b}\bullet R\bullet\underline{a}^T$ follows from the bilinearity of the covriance function
- Let $\underline{\mathbf{Y}} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots \mathbf{Y}_m]$ be obtained by a linear transformation from $\underline{\mathbf{X}}$
- $\underline{\mathbf{Y}} = \underline{\mathbf{X}}\bullet G$ where G is a n×m matrix
- $E[\underline{\mathbf{Y}}] = E[\underline{\mathbf{X}}\bullet G] = E[\underline{\mathbf{X}}]\bullet G = \underline{\mu}_X\bullet G = \underline{\mu}_Y$ since expectation is a linear operation

Covariance of linear combinations

- $\underline{\mathbf{Y}} = \underline{\mathbf{X}}\bullet G$ where G is a n×m matrix
- $\underline{\mathbf{Y}} - \underline{\mu}_Y = \underline{\mathbf{X}}\bullet G - \underline{\mu}_X\bullet G = (\underline{\mathbf{X}} - \underline{\mu}_X)\bullet G$
- $\underline{\mathbf{Y}}$ has a m×m covariance matrix S
- $S = E[(\underline{\mathbf{Y}} - \underline{\mu}_Y)^T\bullet(\underline{\mathbf{Y}} - \underline{\mu}_Y)]$ by definition

  $= E[\{(\underline{\mathbf{X}} - \underline{\mu}_X)\bullet G\}^T\bullet(\underline{\mathbf{X}} - \underline{\mu}_X)\bullet G]$

  $= E[G^T\bullet(\underline{\mathbf{X}} - \underline{\mu}_X)^T\bullet(\underline{\mathbf{X}} - \underline{\mu}_X)\bullet G]$

  $= G^T\bullet E[(\underline{\mathbf{X}} - \underline{\mu}_X)^T\bullet(\underline{\mathbf{X}} - \underline{\mu}_X)]\bullet G$

  $= G^T\bullet R\bullet G$

Summary — I

- The mean of a sum of RVs is the sum of the means of the RVs
- $\text{var}(\mathbf{X}\pm\mathbf{Y}) = \text{var}(\mathbf{X}) + \text{var}(\mathbf{Y}) \pm 2\bullet\text{cov}(\mathbf{X},\mathbf{Y})$
- Variance of a sum of uncorrelated RVs is the sum of the variances of the RVs
- Independent RVs are uncorrelated but uncorrelated RVs are not necessarily independent

Summary — II

- $|\text{cov}(\mathbf{X}, \mathbf{Y})| \leq \sigma_X\bullet\sigma_Y$
- $\rho = \text{cov}(\mathbf{X}, \mathbf{Y})/(\sigma_X\bullet\sigma_Y)$; $|\rho| \leq 1$
- If $\rho = \pm1$, RVs are perfectly correlated and probability mass lies on a straight line (of positive or negative slope) through (μ_X, μ_Y)
- Matrix methods are very helpful in dealing with multiple random variables