Probability 2 - Notes 10

Some Useful Inequalities.

Lemma. If X is a random variable and $g(x) \geq 0$ for all x in the support of f_X, then $P(g(X) \geq 1) \leq E[g(X)]$.

Proof. (continuous case)

$$P(g(X) \geq 1) = \int_{x:g(x)\geq 1} f_X(x)dx \leq \int_{x:g(x)\geq 1} g(x)f_X(x)dx \leq \int_{-\infty}^{\infty} g(x)f_X(x)dx = E[g(X)]$$

Corollaries

1. **Markov's Inequality.** For any $h > 0$, $P(|X| \geq h) \leq \frac{E[|X|]}{h}$. When X only takes non-negative values then for any $h > 0$ $P(X \geq h) \leq \frac{E[X]}{h}$.

Proof. Take $g(X) = \frac{|X|}{h}$ in the lemma. If X only takes non-negative values take $g(X) = \frac{X}{h}$ in the lemma.

2. **Chebyshev's Inequality.** If $E[X] = \mu$ and $Var(X) = \sigma^2$, which are finite, then for any $h > 0$ $P(|X - \mu| \geq h) \leq \frac{\sigma^2}{h^2}$.

Proof. Take $g(X) = \left(\frac{X-\mu}{h}\right)^2$ in the lemma.

Note: Chebyshev's inequality can be used to derive the weak law of large numbers. This is specified in the theorem below.

Theorem. Let $X_1, X_2, ...$ be a sequence of i.i.d. random variables each with finite mean μ and finite variance σ^2. Then for any $\varepsilon > 0$ and $\delta > 0$ there exists an N such that $P(|\overline{X}_n - \mu| \geq \varepsilon) \leq \delta$ for all $n \geq N$, where $\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$.

Proof. Note that $E[\overline{X}_n] = \mu$ and $Var(\overline{X}_n) = \frac{\sigma^2}{n}$. Consider any $\varepsilon > 0$ and $\delta > 0$. Apply Chebyshev's inequality to $\overline{X}_n$ and let $h = \varepsilon$. Then $P(|\overline{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \leq \delta$ provided $n \geq \frac{\sigma^2}{\varepsilon^2\delta}$. Therefore we need only choose $N = \frac{\sigma^2}{\varepsilon^2\delta}$ to obtain the result.

Note. Observe that $\lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \varepsilon) = 0$ for any $\varepsilon > 0$. We say that $\overline{X}_n$ **converges in probability** to μ as n tends to infinity.

Some examples using the inequalities.

1. From Markov's inequality with $h = N\mu$, if X is a non-negative random variable, $P(X > N\mu) \leq \frac{\mu}{N\mu} = \frac{1}{N}$ for any $N > 0$.

2. If $\sigma^2 = 0$ then from Chebyshev's inequality for any $h > 0$, $P(|X - \mu| < h) = 1 - P(|X - \mu| \geq h) \geq 1 - \frac{\sigma^2}{h^2} = 1$. Hence $P(X = \mu) = \lim_{h\downarrow 0} P(|X - \mu| < h) = 1$. So variance zero implies the random variable takes a single value with probability 1.

3. When $\sigma^2 > 0$ Chebyshev's inequality gives a lower bound on the probability that X lies within k standard deviations from the mean. Take $h = k\sigma$. Then

$$P(|X-\mu| < k\sigma) = 1 - P(|X-\mu| \geq k\sigma) \geq 1 - \frac{\sigma^2}{(k\sigma)^2} = 1 - \frac{1}{k^2}$$

4. When $\sigma = 1$, how large a sample is needed if we want to be at least 95% certain that the sample mean lies within 0.5 of the true mean? We use Chebyshev's inequality for $\overline{X}_n$ with $h = 0.5$. Then

$$P(|\overline{X}_n - \mu| < 0.5) = 1 - (|\overline{X}_n - \mu| \geq 0.5) \geq 1 - \frac{\sigma^2}{n(0.5)^2} = 1 - \frac{4}{n} \geq 0.95$$

provided $n \geq \frac{4}{0.05} = 80$. So we need a minimum sample size of 80.

The Central Limit Theorem.

Let $X_1, X_2, ...$ be a sequence of i.i.d. random variables each with finite mean μ and finite variance σ^2 and let $\overline{X}_n$ be the sample mean based on $X_1, ..., X_n$. Then we can find an approximation for $P(\overline{X}_n \leq A)$ when n is large by writing the event for X_n in terms of the standardized variable $Z_n = \sqrt{n}(X_n - \mu)/\sigma$ (i.e. $P(\overline{X}_n \leq A) = P\left(Z_n \leq \frac{\sqrt{n}(A-\mu)}{\sigma}\right)$) and proving that $\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ which is the c.d.f. of $N(0,1)$. The proof of this result uses the m.g.f. and the following lemma.

Lemma. Let $Z_1, Z_2, ...$ be a sequence of random variables. If $\lim_{n\to\infty} M_{Z_n}(t) = M(t)$, which is the m.g.f. of a distribution with c.d.f. F, then $\lim_{n\to\infty} F_{Z_n}(z) = F(z)$ at all points z for which $F(z)$ is continuous.

Theorem (The Central Limit Theorem). Let $X_1, X_2, ...$ be a sequence of i.i.d. random variables each with m.g.f. which exists for entries in an open region about zero so is differentiable with finite mean, denoted by μ, and finite variance, denoted by σ^2. Let $Z_n = \sqrt{n}(\overline{X}_n - \mu)/\sigma$, then $\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z)$.

Proof. Let $U_j = (X_j - \mu)/\sigma$ and let $M_U(t)$ be the common m.g.f. Then $M_U(t) = e^{-\mu t/\sigma} M_X(t/\sigma)$ exists in an open interval about $t = 0$, $M(0) = 1$, $M'(0) = E[U] = 0$ and $M''(0) = E[U^2] = Var(U) = 1$. So $U_1, U_2, ...$ are i.i.d. with mean zero and variance one. Now

$$M_{Z_n}(t) = E\left[e^{t\sum_{j=1}^{n} U_j/\sqrt{n}}\right] = \prod_{j=1}^{n} E[e^{tU_j/\sqrt{n}}] = \left(M_U(t/\sqrt{n})\right)^n$$

Taking logs to base e gives $ln(M_{Z_n}(t)) = n(ln(M_U(t/\sqrt{n})))$. Now let $x = 1/\sqrt{n}$ and use L'Hopital's rule. Then

$$\lim_{n\to\infty} n \ln(M_U(t/\sqrt{n})) = \lim_{x\downarrow 0} \frac{ln(M_U(xt))}{x^2}$$

$$= \lim_{x\downarrow 0} \frac{tM_U'(xt)/M_U(xt)}{2x} = \lim_{x\downarrow 0} \frac{t^2(M_U''(xt)M_U(xt) - (M_U'(xt))^2)/(M_U(xt))^2}{2}$$
$$= \frac{t^2(M_U''(0)M_U(0) - (M_U'(0))^2)}{2(M_U(0))^2} = \frac{t^2}{2}$$

Hence $\lim_{t\to\infty} ln(M_{Z_n}(t)) = t^2/2$ and so $\lim_{t\to\infty} M_{Z_n}(t) = e^{t^2/2}$. Since this is the m.g.f. of an $N(0,1)$ distribution, using the lemma proves that $\lim_{n\to\infty} P(Z_n \le z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$

The bivariate and multivariate normal distribution.

An indirect method was used on problem sheet 9 to get you to derive standard results for a bivariate normal distribution. The results are summarised below. The results may be proved directly, however it is messy unless you use matrix and vector notation. Once you do this results can just as easily be obtained for the multivariate normal, so we may just as well derive results immediately for the more general case.

Summary of results for the bivariate normal distribution.

1. If X_1 and X_2 have bivariate normal distribution then the joint p.d.f. is

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} e^{\frac{-1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)}$$

for all x, y. The distribution has parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$. The parameter ρ is restricted so that $-1 < \rho < 1$.

2. The joint m.g.f. is

$$M_{X_1,X_2}(t_1,t_2) = e^{(\mu_1 t_1 + \mu_2 t_2) + \frac{1}{2}(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2)}$$

This can be used to identify the parameters and find the marginal distributions. $M_{X_1}(t_1) = M_{X_1,X_2}(t_1,0) = e^{\mu_1 t_1 + \frac{1}{2}\sigma_1^2 t_1^2}$. Hence $X_1 \sim N(\mu_1, \sigma_1^2)$. Similarly $X_2 \sim N(\mu_2, \sigma_2^2)$. Differentiating the joint m.g.f. in standard manner shows that $\rho(X_1, X_2) = \rho$.

3. X_1 and X_2 are independent iff $\rho = 0$. This is easily seen from either the joint p.d.f. or the joint m.g.f.

4. The conditional distribution of $X_2|X_1 = x_1$ is normal with mean linear in x_1 and variance which does not depend on x_1. A similar result holds for the conditional distribution of $X_1|X_2 = x_2$.

Using vector and matrix notation.

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}; \; \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \; \mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}; \; \mathbf{m} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \; \mathbf{V} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Then $\mathbf{m}$ is the vector of means and $\mathbf{V}$ is the variance-covariance matrix. Note that $|\mathbf{V}| = \sigma_1^2\sigma_2^2(1-\rho^2)$ and

$$\mathbf{V}^{-1} = \frac{1}{(1-\rho^2)}\begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$$

Hence $f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{2/2}|\mathbf{V}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m})}$ for all $\mathbf{x}$. Also $M_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}^T\mathbf{m}+\frac{1}{2}\mathbf{t}^T\mathbf{V}\mathbf{t}}$.

The Multivariate Normal Distribution.

We again use matrix and vector notation, but now there are n random variables so that $\mathbf{X}$, $\mathbf{x}$, $\mathbf{t}$ and $\mathbf{m}$ are now n-vectors with i^{th} entries X_i, x_i, t_i and μ_i and $\mathbf{V}$ is the $n \times n$ matrix with ii^{th} entry σ_i^2 and ij^{th} entry (for $i \neq j$) σ_{ij}. Note that $\mathbf{V}$ is symmetric so that $\mathbf{V}^T = \mathbf{V}$.

The joint p.d.f. is $f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}|\mathbf{V}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m})}$ for all $\mathbf{x}$. We say that $\mathbf{X} \sim N(\mathbf{m}, \mathbf{V})$.

We can find the joint m.g.f. quite easily.

$$M_{\mathbf{X}}(\mathbf{t}) = E\left[e^{\sum_{j=1}^n t_j X_j}\right] = E[e^{\mathbf{t}^T\mathbf{X}}] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2}|\mathbf{V}|^{1/2}} e^{-\frac{1}{2}\left((\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m})-2\mathbf{t}^T\mathbf{x}\right)} dx_1 \ldots dx_n$$

We do the equivalent of completing the square, i.e. we write

$$(\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}) - 2\mathbf{t}^T\mathbf{x} = (\mathbf{x}-\mathbf{m}-\mathbf{a})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}-\mathbf{a}) + b$$

for a suitable choice of the n-vector $\mathbf{a}$ of constants and a constant b. Then

$$M_{\mathbf{X}}(\mathbf{t}) = e^{-b/2}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{n/2}|\mathbf{V}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}-\mathbf{a})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}-\mathbf{a})} dx_1 \ldots dx_n = e^{-b/2}.$$

We just need to find $\mathbf{a}$ and b. Expanding we have

$$\begin{aligned} & ((\mathbf{x}-\mathbf{m})-\mathbf{a})^T\mathbf{V}^{-1}((\mathbf{x}-\mathbf{m})-\mathbf{a}) + b \\ = \; & (\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}) - 2\mathbf{a}^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}) + \mathbf{a}^T\mathbf{V}^{-1}\mathbf{a} + b \\ = \; & (\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}) - 2\mathbf{a}^T\mathbf{V}^{-1}\mathbf{x} + \left[2\mathbf{a}^T\mathbf{V}^{-1}\mathbf{m} + \mathbf{a}^T\mathbf{V}^{-1}\mathbf{a} + b\right] \end{aligned}$$

This has to equal $(\mathbf{x}-\mathbf{m})^T\mathbf{V}^{-1}(\mathbf{x}-\mathbf{m}) - 2\mathbf{t}^T\mathbf{x}$ for all $\mathbf{x}$. Hence we need $\mathbf{a}^T\mathbf{V}^{-1} = \mathbf{t}^T$ and $b = -\left[2\mathbf{a}^T\mathbf{V}^{-1}\mathbf{m} + \mathbf{a}^T\mathbf{V}^{-1}\mathbf{a}\right]$. Hence $\mathbf{a} = \mathbf{V}\mathbf{t}$ and $b = -\left[2\mathbf{t}^T\mathbf{m} + \mathbf{t}^T\mathbf{V}\mathbf{t}\right]$. Therefore

$$M_{\mathbf{X}}(\mathbf{t}) = e^{-b/2} = e^{\mathbf{t}^T\mathbf{m}+\frac{1}{2}\mathbf{t}^T\mathbf{V}\mathbf{t}}$$

Results obtained using the m.g.f.

1. Any (non-empty) subset of multivariate normals is multivariate normal. Simply put $t_j = 0$ for all j for which X_j is not in the subset. For example $M_{X_1}(t_1) = M_{X_1,...,X_n}(t_1, 0, ..., 0) = e^{t_1\mu_1 + t_1^2\sigma_1^2/2}$. Hence $X_1 \sim N(\mu_1, \sigma_1^2)$. A similar result holds for X_i. This identifies the parameters μ_i and σ_i^2 as the mean and variance of X_i. Also

$$M_{X_1,X_2}(t_1,t_2) = M_{X_1,...,X_n}(t_1,t_2,0,...,0) = e^{t_1\mu_1+t_2\mu_2+\frac{1}{2}(t_1^2\sigma_1^2+2\sigma_{12}t_1t_2+\sigma_2^2t_2^2)}$$

Hence X_1 and X_2 have bivariate normal distribution with $\sigma_{12} = Cov(X_1, X_2)$. A similar result holds for the joint distribution of X_i and X_j for $i \neq j$. This identifies $\mathbf{V}$ as the variance-covariance matrix for $X_1, ..., X_n$.

2. $\mathbf{X}$ is a vector of independent random variables iff $\mathbf{V}$ is diagonal (i.e. all off-diagonal entries are zero so that $\sigma_{ij} = 0$ for $i \neq j$).

Proof. From (1), if the $X's$ are independent then $\sigma_{ij} = Cov(X_i, X_j) = 0$ for all $i \neq j$, so that $\mathbf{V}$ is diagonal.

If $\mathbf{V}$ is diagonal then $\mathbf{t}^T\mathbf{V}\mathbf{t} = \sum_{j=1}^{n} \sigma_j^2 t_j^2$ and hence

$$M_{\mathbf{X}}(\mathbf{t}) = e^{\mathbf{t}^T\mathbf{m}+\frac{1}{2}\mathbf{t}^T\mathbf{V}\mathbf{t}} = \prod_{j=1}^{n}\left(e^{\mu_j t_j+\frac{1}{2}\sigma_j^2 t_j^2/2}\right) = \prod_{j=1}^{n} M_{X_j}(t_j)$$

By the uniqueness of the joint m.g.f., $X_1, ..., X_n$ are independent.

3. Linearly independent linear functions of multivariate normal random variables are multivariate normal random variables. If $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, where $\mathbf{A}$ is an $n \times n$ non-singular matrix and $\mathbf{b}$ is a (column) n-vector of constants, then $\mathbf{Y} \sim N(\mathbf{A}\mathbf{m} + \mathbf{b}, \mathbf{A}\mathbf{V}\mathbf{A}^T)$.

Proof. Use the joint m.g.f.

$$\begin{aligned} M_{\mathbf{Y}}(\mathbf{t}) &= E[e^{\mathbf{t}^T\mathbf{Y}}] = E[e^{\mathbf{t}^T\mathbf{A}\mathbf{X}+\mathbf{b}}] = e^{\mathbf{t}^T\mathbf{b}}E[e^{(\mathbf{A}^T\mathbf{t})^T\mathbf{X}}] = e^{\mathbf{t}^T\mathbf{b}}M_{\mathbf{X}}(\mathbf{A}^T\mathbf{t}) \\ &= e^{\mathbf{t}^T\mathbf{b}}e^{(\mathbf{A}^T\mathbf{t})^T\mathbf{m}+\frac{1}{2}(\mathbf{A}^T\mathbf{t})^T\mathbf{V}(\mathbf{A}^T\mathbf{t})} = e^{\mathbf{t}^T(\mathbf{A}\mathbf{m}+\mathbf{b})+\frac{1}{2}\mathbf{t}^T(\mathbf{A}\mathbf{V}\mathbf{A}^T)\mathbf{t}} \end{aligned}$$

This is just the m.g.f. for the multivariate normal distribution with vector of means $\mathbf{A}\mathbf{m} + \mathbf{b}$ and variance-covariance matrix $\mathbf{A}\mathbf{V}\mathbf{A}^T$. Hence, from the uniqueness of the joint m.g.f, $\mathbf{Y} \sim N(\mathbf{A}\mathbf{m} + \mathbf{b}, \mathbf{A}\mathbf{V}\mathbf{A}^T)$.

Note that from (2) a subset of the $Y's$ is multivariate normal.

NOTE. The results concerning the vector of means and variance-covariance matrix for linear functions of random variables hold regardless of the joint distribution of $X_1, ..., X_n$.

We define the expectation of a vector of random variables $\mathbf{X}$, $E[\mathbf{X}]$ to be the vector of the expectations and the expectation of a matrix of random variables $\mathbf{Y}$, $E[\mathbf{Y}]$, to be the matrix of the expectations. Then the variance-covariance matrix of $\mathbf{X}$ is just $E[(\mathbf{X}-E[\mathbf{X}])(\mathbf{X}-E[\mathbf{X}])^T]$.

The following results are easily obtained:

(i) Let $\mathbf{A}$ be an $m \times n$ matrix of constants, $\mathbf{B}$ be an $m \times k$ matrix of constants and $\mathbf{Y}$ be an $n \times k$ matrix of random variables. Then $E[\mathbf{AY}+\mathbf{B}] = \mathbf{A}E[\mathbf{Y}]+\mathbf{B}$.

Proof. The ij^{th} entry of $E[\mathbf{AY}+\mathbf{B}]$ is $E[\sum_{r=1}^{n} A_{ir}Y_{rj}+B_{ij}] = \sum_{r=1}^{n} A_{ir}E[Y_{rj}]+B_{ij}$, which is the ij^{th} entry of $\mathbf{A}E[\mathbf{Y}]+\mathbf{B}$. The result is then immediate.

(ii) Let $\mathbf{C}$ be a $k \times m$ matrix of constants and $\mathbf{Y}$ be an $n \times k$ matrix of random variables. Then $E[\mathbf{YC}] = E[\mathbf{Y}]\mathbf{C}$.

Proof. Just transpose the equation. The result then follows from (i).

Hence if $\mathbf{Z} = \mathbf{AX}+\mathbf{b}$, where $\mathbf{A}$ is an $m \times n$ matrix of constants, $\mathbf{b}$ is an m-vector of constants and $\mathbf{X}$ is an n-vector of random variables with $E[\mathbf{X}] = \mathbf{m}$ and variance-covariance matrix $\mathbf{V}$, then

$$E[\mathbf{Z}] = E[\mathbf{AX}+\mathbf{b}] = \mathbf{A}E[\mathbf{X}]+\mathbf{b} = \mathbf{Am}+\mathbf{b}$$

Also the variance-covariance matrix for $\mathbf{Y}$ is just

$$E[(\mathbf{Y}-E[\mathbf{Y}])(\mathbf{Y}-E[\mathbf{Y}])^T] = E[\mathbf{A}(\mathbf{X}-\mathbf{m})(\mathbf{X}-\mathbf{m})^T\mathbf{A}^T] = \mathbf{A}E[(\mathbf{X}-\mathbf{m})(\mathbf{X}-\mathbf{m})^T]\mathbf{A}^T = \mathbf{AVA}^T$$

Example. Suppose that $E[X_1] = 1$, $E[X_2] = 0$, $Var(X_1) = 2$, $Var(X_2) = 4$ and $Cov(X_1,X_2) = 1$. Let $Y_1 = X_1+X_2$ and $Y_2 = X_1+aX_2$. Find the means, variances and covariance and hence find a so that Y_1 and Y_2 are uncorrelated.

Writing in vector and matrix notation we have $E[\mathbf{Y}] = \mathbf{Am}$ and the variance-covariance matrix for $\mathbf{Y}$ is just $\mathbf{AVA}^T$ where

$$\mathbf{m} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \mathbf{V} = \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \quad \mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & a \end{pmatrix}$$

Therefore

$$\mathbf{Am} = \begin{pmatrix} 1 & 1 \\ 1 & a \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\mathbf{AVA}^T = \begin{pmatrix} 1 & 1 \\ 1 & a \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & a \end{pmatrix} = \begin{pmatrix} 8 & 3+5a \\ 3+5a & 2+2a+4a^2 \end{pmatrix}$$

Hence Y_1 and Y_2 have means 1 and 1, variances 8 and $2+2a+4a^2$ and covariance $3+5a$. They are therefore uncorrelated if $3+5a=0$, i.e. if $a=-\frac{3}{5}$.