Bivariate dependence

Probability I (BST 230)

Jeffrey W. Miller

Department of Biostatistics
Harvard T.H. Chan School of Public Health

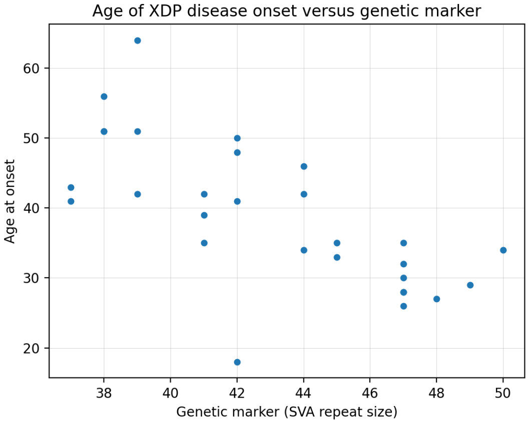
Outline

Outline

Correlation and covariance

- When random variables X and Y are not independent, they are dependent.
- However, the dependence may be weak, or it may be strong.
- Correlation is an important way of quantifying the dependence between random variables.
- Covariance is a related concept that also depends on the scales of X and Y.

Example: Age of disease onset versus a genetic marker



Correlation and covariance

- To declutter the notation, let's denote:

$$\mu_X = \mathrm{E}X \qquad \sigma_X = \sqrt{\operatorname{Var} X}$$
$$\mu_Y = \mathrm{E}Y \qquad \sigma_Y = \sqrt{\operatorname{Var} Y}.$$

- The *covariance* of X and Y is

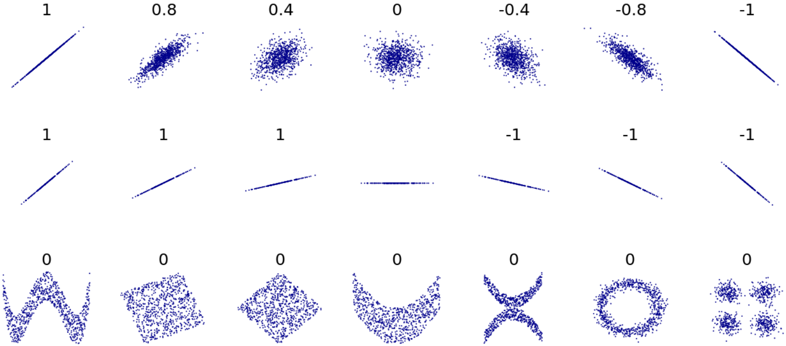
$$\operatorname{Cov}(X, Y) = \mathrm{E}\Big((X - \mu_X)(Y - \mu_Y)\Big).$$

- The *correlation* of X and Y is

$$\operatorname{Corr}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

- The correlation is often denoted $\rho_{X,Y}$, and also called the "Pearson correlation" or the "correlation coefficient".

Examples of data with various correlations



Properties of correlation and covariance

- $-1 \leq \rho_{X,Y} \leq 1$.
- $\rho_{X,Y} > 0$ implies a positive association (direct relationship); $\rho_{X,Y} < 0$ implies a negative association (inverse relationship).
- If $\sigma_X = 0$ or $\sigma_Y = 0$, then $\rho_{X,Y}$ is undefined.
- $|\rho_{X,Y}| = 1$ if and only if there exist $a \neq 0$ and $b \in \mathbb{R}$ such that $P(Y = aX + b) = 1$. The sign of $\rho_{X,Y}$ equals the sign of a.
- Correlation captures the strength of association in terms of how close to linear it is, but not the magnitude of the slope.
- If X and Y are independent, then $\rho_{X,Y} = 0$. However, if $\rho_{X,Y} = 0$ then X and Y are not necessarily independent.

Properties of covariance

1. $\mathrm{Cov}(X,Y) = \rho_{X,Y}\sigma_X\sigma_Y$.

2. $\mathrm{Cov}(X,Y) = \mathrm{E}XY - (\mathrm{E}X)(\mathrm{E}Y)$.

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

4. $\mathrm{Cov}(aX+b,\, cY+d) = a\,c\,\mathrm{Cov}(X,Y)$.

5. $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$.

Group exercise (10 minutes): Show 2, 3, and 5.

Outline

Bivariate transformations

- Suppose (X, Y) is a random vector, and $(U, V) = g(X, Y)$ for some function g.
- That is, $U = g_1(X, Y)$ and $V = g_2(X, Y)$.
- How can we derive the joint pdf/pmf of (U, V) from the joint pdf/pmf of (X, Y)?
- We saw how to do this in the univariate case. Now we extend to the bivariate case.

Bivariate transformations: Discrete case

- Suppose $(U, V) = g(X, Y)$ for some function g.

- If (X, Y) is discrete, then the joint pmf of (U, V) is

$$f_{U,V}(u, v) = \sum_{x,y} f_{X,Y}(x, y) \mathbb{1}\big(g(x, y) = (u, v)\big).$$

- This is really just the same as the univariate case, except that we are considering bivariate rather than univariate elements.

Bivariate transformations: Continuous case

- It is much less obvious how to handle the case where (X, Y) is continuous. Fortunately, however, there is still a nice formula.

- Suppose (X, Y) is a continuous random vector, and $(U, V) = g(X, Y)$ for some function g such that:
  1. g is one-to-one, with inverse $h(u, v) = (x, y)$ on its range,
  2. the partial derivatives of $g(x, y)$ exist and are continuous,
  3. the Jacobian matrix Dh is nonsingular, where

$$Dh = \begin{bmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial h_1}{\partial u} & \dfrac{\partial h_1}{\partial v} \\ \\ \dfrac{\partial h_2}{\partial u} & \dfrac{\partial h_2}{\partial v} \end{bmatrix}.$$

  The joint pdf of (U, V) is

$$f_{U,V}(u, v) = f_{X,Y}\big(h_1(u, v), h_2(u, v)\big)\big|\det(Dh)\big|$$

  for (u, v) in the range of $g(X, Y)$, and is zero elsewhere.

Bivariate transformations: Details

- $g(x, y)$ being one-to-one means that if $(x, y) \neq (x', y')$ then $g(x, y) \neq g(x', y')$. A one-to-one function always has a inverse from its range back to its domain.

- The determinant factor is

$$\big|\det(Dh)\big| = \left|\frac{\partial h_1}{\partial u}\frac{\partial h_2}{\partial v} - \frac{\partial h_1}{\partial v}\frac{\partial h_2}{\partial u}\right|.$$

- Sometimes $\det(Dh)$ is referred to as the *Jacobian* of h, often denoted J_h or simply J. The notation is not totally standard though, and J_h sometimes denotes the matrix Dh.

Example: A transformation of standard normals

- Suppose $X, Y \sim \mathcal{N}(0,1)$ independently, and $\rho \in (-1,1)$.

- Define $U = X$ and $V = \rho X + \sqrt{1-\rho^2}Y$. That is,
$$
\begin{aligned}
u &= g_1(x,y) = x \\
v &= g_2(x,y) = \rho x + \sqrt{1-\rho^2}\, y.
\end{aligned}
$$

- The inverse is defined by
$$
\begin{aligned}
x &= h_1(u,v) = u \\
y &= h_2(u,v) = \frac{v-\rho u}{\sqrt{1-\rho^2}}.
\end{aligned}
$$

- Thus, the Jacobian matrix is
$$
Dh = \begin{bmatrix} \dfrac{\partial h_1}{\partial u} & \dfrac{\partial h_1}{\partial v} \\ \dfrac{\partial h_2}{\partial u} & \dfrac{\partial h_2}{\partial v} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\dfrac{\rho}{\sqrt{1-\rho^2}} & \dfrac{1}{\sqrt{1-\rho^2}} \end{bmatrix}.
$$

Example: A transformation of standard normals

- Therefore, the joint pdf of (U, V) is

$$
\begin{aligned}
& f_{U,V}(u,v) = f_{X,Y}(h_1(u,v), h_2(u,v)) \big| \det(Dh) \big| \\
& = \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}u^2) \frac{1}{\sqrt{2\pi}} \exp\Big(- \frac{1}{2}\Big(\frac{v-\rho u}{\sqrt{1-\rho^2}}\Big)^2\Big) \Big| \frac{1}{\sqrt{1-\rho^2}} \Big| \\
& = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big(- \frac{1}{2}u^2 - \frac{1}{2}\Big(\frac{v^2 - 2v\rho u + \rho^2 u^2}{1-\rho^2}\Big)\Big) \\
& = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\Big(- \frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\Big).
\end{aligned}
$$

- This is the pdf of a *bivariate normal distribution* with *correlation* ρ, mean zero, and unit variances.

- More generally, a bivariate normal distribution can have any mean and variances.

Example: A transformation of uniforms

- Let $X, Y \sim \mathrm{Uniform}(-1, 1)$ independently.
- Suppose $U = (X + Y)/2$ and $V = X - Y$.

  Group exercise (10 mins): What is the joint pdf of (U, V)? Also, draw a picture of the joint pdf.

Transforming two variables into one variable

- Often, we want to know the distribution of a single random variable $U = g_1(X, Y)$ that is a function of (X, Y).
- However, this is hardly ever an invertible transformation.
- Fortunately, it turns out that we can still use the bivariate transformation technique, as follows.
- Introduce a new "auxiliary" variable $V = g_2(X, Y)$, chosen to make calculations as easy as possible.
- Compute $f_{U,V}(u, v)$ from $f_{X,Y}(x, y)$ using the bivariate transformation formula.
- Then, integrate to get the marginal density of U:

$$f_U(u) = \int f_{U,V}(u, v)\mathrm{d}v.$$

Example: Ratio of standard normals

- To illustrate this technique, suppose $X, Y \sim \mathcal{N}(0,1)$ independently. What is the distribution of X/Y?

- Define $U = X/Y$ and $V = Y$. That is,

$$u = g_1(x,y) = x/y$$
$$v = g_2(x,y) = y.$$

- Introducing $V = Y$ makes g invertible, so we can use the bivariate transformation formula.

- The inverse is $x = h_1(u,v) = uv$ and $y = h_2(u,v) = v$.

- Thus, the Jacobian matrix is

$$Dh = \begin{bmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{bmatrix} = \begin{bmatrix} v & u \\ 0 & 1 \end{bmatrix}.$$

Example: Ratio of standard normals

- Therefore, the joint pdf of (U, V) is

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}(h_1(u,v), h_2(u,v))\big|\det(Dh)\big| \\ &= \frac{1}{\sqrt{2\pi}}\exp\big(-\tfrac{1}{2}(uv)^2\big)\frac{1}{\sqrt{2\pi}}\exp\big(-\tfrac{1}{2}v^2\big)|v| \\ &= \frac{1}{2\pi}|v|\exp\big(-\tfrac{1}{2}v^2(1+u^2)\big). \end{aligned}$$

- To find the marginal of U, we need to integrate out V.
- Making the change of variable $t = v^2$, $\mathrm{d}t = 2v\mathrm{d}v$,

$$\begin{aligned} \int_0^\infty v\exp\big(-\tfrac{1}{2}v^2(1+u^2)\big)\mathrm{d}v &= \frac{1}{2}\int_0^\infty \exp\big(-\tfrac{1}{2}t(1+u^2)\big)\mathrm{d}t \\ &= \frac{1}{1+u^2}. \end{aligned}$$

- Therefore, since $f_{U,V}(u,-v) = f_{U,V}(u,v)$,

$$f_U(u) = \int_{-\infty}^\infty f_{U,V}(u,v)\mathrm{d}v = 2\int_0^\infty f_{U,V}(u,v)\mathrm{d}v = \frac{1}{\pi}\frac{1}{1+u^2}.$$

- Do you recognize this distribution?

Outline

Bivariate normal distribution

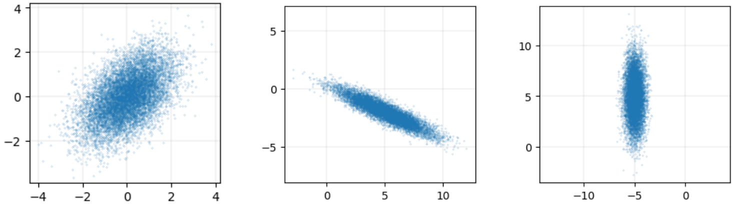
- The *bivariate normal* distribution with means $\mu_X, \mu_Y \in \mathbb{R}$, variances $\sigma_X^2, \sigma_Y^2 > 0$, and correlation $\rho \in (-1, 1)$ has pdf

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\tilde{x}^2 - 2\rho\tilde{x}\tilde{y} + \tilde{y}^2\right)\right)$$

  where

$$\tilde{x} = \frac{x-\mu_X}{\sigma_X} \quad \text{and} \quad \tilde{y} = \frac{y-\mu_Y}{\sigma_Y}.$$

- Examples with varying means, variances, and correlations:



Bivariate normal distribution: Properties

- If (X, Y) follow a bivariate normal distribution, then:

1. $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.

2. $\mathrm{Corr}(X, Y) = \rho$.

3. $aX + bY$ is normally distributed for any $a, b \in \mathbb{R}$.

   For the $a = b = 0$ case, we need to expand our definition of the family of normal distributions to allow $\sigma^2 = 0$ by considering $\mathcal{N}(\mu, 0)$ to be the point mass at μ.

4. $Y \mid X = x$ is normally distributed with

$$\mathrm{E}(Y \mid X = x) = \mu_Y + \rho\sigma_Y \frac{x - \mu_X}{\sigma_X}$$
$$\mathrm{Var}(Y \mid X = x) = (1 - \rho^2)\sigma_Y^2.$$

   In other words, the conditional pdf of Y given $X = x$ is

$$p(y|x) = \mathcal{N}\left(y \,\middle|\, \mu_Y + \rho\sigma_Y \frac{x - \mu_X}{\sigma_X},\ (1 - \rho^2)\sigma_Y^2\right).$$

Bivariate normal distribution: Properties

- If X and Y are each normally distributed, then (X, Y) is NOT necessarily bivariate normal.
  - **Group exercise (3 minutes)**: Can you come up with an example to illustrate this?

- We defined the bivariate normal distribution in terms of its pdf, but there is a more general definition that we will use.

- *Definition:* We say that (X, Y) is *bivariate normal* if $aX + bY$ is normally distributed for all $a, b \in \mathbb{R}$.

- How is this definition more general? That is, when does this definition apply but the pdf definition doesn't apply?

Mean and covariance of a random vector

- The *covariance matrix* of a random vector $(X, Y)^{\mathrm{T}} = \begin{bmatrix} X \\ Y \end{bmatrix}$ is

$$\mathrm{Cov}\left(\begin{bmatrix} X \\ Y \end{bmatrix} \right) = \begin{bmatrix} \mathrm{Cov}(X,X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Cov}(Y,Y) \end{bmatrix}$$
$$= \begin{bmatrix} \sigma_X^2 & \rho_{X,Y}\sigma_X\sigma_Y \\ \rho_{X,Y}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}.$$

- The mean of a random vector $(X, Y)^{\mathrm{T}}$ is defined to be the vector of the means of its entries:

$$\mathrm{E}\left(\begin{bmatrix} X \\ Y \end{bmatrix} \right) = \begin{bmatrix} \mathrm{E}X \\ \mathrm{E}Y \end{bmatrix}.$$

- For any random vector $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ and any 2×2 matrix A,

$$\mathrm{Cov}(AX) = A\,\mathrm{Cov}(X)A^{\mathtt{T}}.$$

(This holds more generally in the d-dimensional case.)

Bivariate normal: Mean and covariance parametrization

- It is common to parametrize bivariate (and more generally, multivariate) normal distributions in terms of the mean vector and covariance matrix. We write

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$$

to denote that $(X, Y)^{\mathrm{T}}$ is bivariate normal such that

$$\mu = \mathrm{E}\Big(\begin{bmatrix} X \\ Y \end{bmatrix} \Big) \quad \text{and} \quad \Sigma = \mathrm{Cov}\Big(\begin{bmatrix} X \\ Y \end{bmatrix} \Big).$$

- However, not just any 2×2 matrix Σ can be used. Σ must be a *symmetric positive semi-definite* matrix, that is,

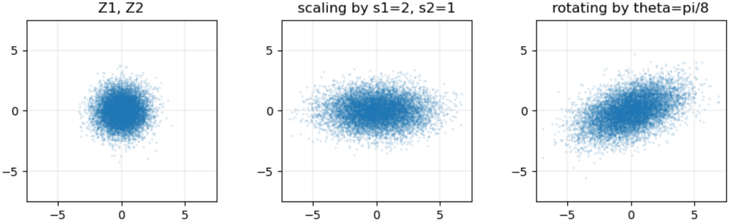
  1. $\Sigma = \Sigma^{\mathtt{T}}$ (symmetric), and
  2. $t^{\mathtt{T}} \Sigma t \geq 0$ for all $t \in \mathbb{R}^2$ (positive semi-definite).

Bivariate normal: Scale/rotation construction

- This leads to a useful way of constructing bivariate normals.
- Let $s_1 \geq s_2 \geq 0$ and $\theta \in [0, 2\pi)$.
- Let $Z_1, Z_2 \sim \mathcal{N}(0,1)$ independently, and define

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \underbrace{\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}}_{\text{rotation}} \underbrace{\begin{bmatrix} s_1 & 0 \\ 0 & s_2 \end{bmatrix}}_{\text{scaling}} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}.$$

- Then $(X_1, X_2)^{\mathrm{T}}$ is bivariate normal such that the line along which X_1 and X_2 are correlated is at angle θ, the scale along this line is s_1, and the scale orthogonal to the line is s_2.



Bivariate normal: Scale/rotation decomposition

- Conversely, given Σ, we can recover the scaling and rotation.
- Compute the "eigendecomposition" $\Sigma = U\Lambda U^{\mathtt{T}}$ where U is an orthogonal matrix and $\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ with $\lambda_1 \geq \lambda_2 \geq 0$.
  - A matrix is *orthogonal* if $U^{\mathtt{T}}U = I$ and $UU^{\mathtt{T}} = I$.
- Then $\lambda_1 = s_1^2$, $\lambda_2 = s_2^2$, and U is the rotation matrix.
- Then, we can represent $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}(0, \Sigma)$ as

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \underbrace{\begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}}_{U \text{ (rotation)}} \underbrace{\begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix}}_{\Lambda^{1/2} \text{ (scaling)}} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

  where $Z_1, Z_2 \sim \mathcal{N}(0, 1)$ independently.
- Or, more succinctly, $X = U\Lambda^{1/2}Z$.

Bivariate normal: Scale/rotation decomposition

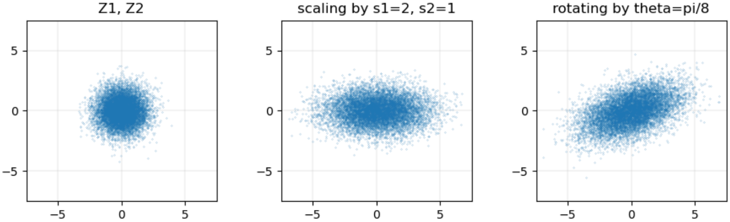
- But starting from $\Sigma = U \Lambda U^{\mathtt{T}}$ and $Z = (Z_1, Z_2)^{\mathtt{T}}$ in this way, how can we be sure that $X = U \Lambda^{1/2} Z$ is bivariate normal?

  Group exercise (5 minutes): See if you can figure out why! (Hint: Use the general definition, not the pdf definition.)

Connection to principal components analysis (PCA)

- Principal components analysis (PCA) can be done by applying this decomposition to the sample covariance matrix $\hat{\Sigma}$ estimated from data $x_1, \ldots, x_n \in \mathbb{R}^2$.

- The columns of U are the PC directions, s_1, s_2 are the PC scales, $U^{\mathrm{T}} x_i$ are the PC scores.



Bivariate normal: Moment generating function

- The mgf of a bivariate random vector $X = (X_1, X_2)^{\mathtt{T}} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ is defined to be

$$M_X(t) = \mathrm{E}\exp(t_1 X_1 + t_2 X_2) = \mathrm{E}\exp(t^{\mathtt{T}} X)$$

for $t = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \in \mathbb{R}^2$.

- If X is bivariate normal then

$$M_X(t) = \exp\left(t_1\mu_{X_1} + t_2\mu_{X_2} + \frac{1}{2}\left(t_1^2\sigma_{X_1}^2 + 2t_1t_2\rho\sigma_{X_1}\sigma_{X_2} + t_2^2\sigma_{X_2}^2\right)\right).$$

- In matrix/vector notation, if $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$ then

$$M_X(t) = \exp\left(t^{\mathtt{T}}\mu + \frac{1}{2}t^{\mathtt{T}}\Sigma t\right).$$

Group exercise (5 minutes): Can you show this?
(Hint: Use the mgf of a univarite normal.)