Lecture 09: Covariance and Correlation

Mathematical Statistics I, MATH 60061/70061

Tuesday September 28, 2021

Reference: Casella & Berger, 4.5

Covariance

The **covariance** between random variables X and Y is

$$\mathrm{Cov}(X,Y) = E((X - EX)(Y - EY)).$$

Using linearity, we have an equivalent expression:

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E(XY - X(EY) - Y(EX) + (EX)(EY)) \\ &= E(XY) - E(X)E(Y).\end{aligned}$$

We say that random variables with zero covariance are **uncorrelated**.

Uncorrelated vs. independence

If X and Y are independent, then they are uncorrelated.

$$\begin{aligned}
E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)f_Y(y)dxdy \\
&= \int_{-\infty}^{\infty} yf_Y(y)\left(\int_{-\infty}^{\infty} xf_X(x)dx\right)dy \\
&= \int_{-\infty}^{\infty} xf_X(x)dx\int_{-\infty}^{\infty} yf_Y(y)dy \\
&= E(X)E(Y).
\end{aligned}$$

Uncorrelated vs. independence, continued

The fact that X and Y are uncorrelated does not imply that they are independent.

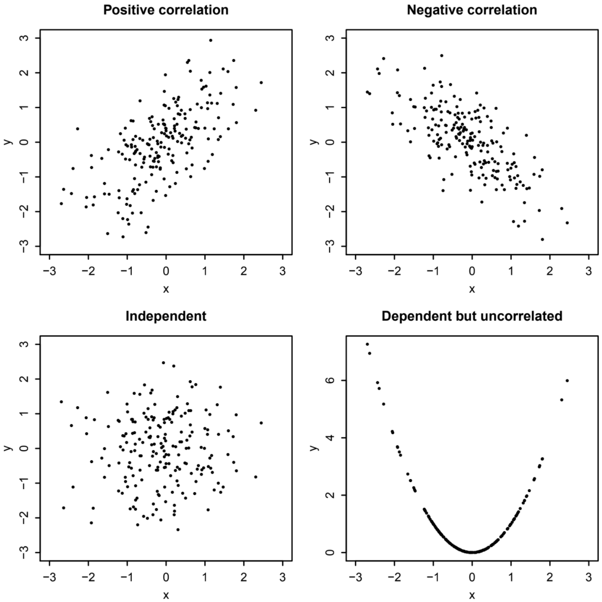
For example, let $X \sim \mathcal{N}(0,1)$, and let $Y = X^2$. Then $E(XY) = E(X^3) = 0$. Thus X and Y are uncorrelated,

$$\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y) = 0 - 0 = 0,$$

but they are *not* independent: knowing X gives perfect information about Y.

Covariance is a measure of *linear* association, so random variables can be dependent in nonlinear ways and still have zero covariance.

Joint distributions under various dependence structures



Key properties of covariance

1. $\text{Cov}(X, X) = \text{Var}(X)$.
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
3. $\text{Cov}(X, c) = 0$ for any constant c.
4. $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$ for any constant a.
5. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$.
6. $\text{Cov}(X + Y, Z + W) =$
   $\text{Cov}(X, Z) + \text{Cov}(X, W) + \text{Cov}(Y, Z) + \text{Cov}(Y, W)$.
7. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$.
8. $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$.
9. For n random variables $X_1, \ldots, X_n$,

$$\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n) + 2\sum_{i<j} \text{Cov}(X_i, X_j).$$

Multinomial distribution

The Multinomial distribution is a generalization of the Binomial.

- Binomial counts the successes in a fixed number of trials that can only be categorized as success or failure.
- Multinomial keeps track of trials whose outcomes can fall into multiple categories.

Each of n objects is independently placed into one of the k categories. An object is placed into category j with probability p_j, where the p_j's are nonnegative and $\sum_{j=1}^{k} p_j = 1$. Let X_1 be the number of objects in category 1, X_2 be the number of objects in category 2, etc., so that $X_1 + \cdots + X_k = n$. Then $\boldsymbol{X} = (X_1, \ldots, X_k)$ is said to have the **Multinomial distribution** with parameters n and $\boldsymbol{p} = (p_1, \ldots, p_k)$, $\boldsymbol{X} \sim \mathrm{Mult}_k(n, \boldsymbol{p})$.

Multinomial joint and marginal distributions

If $\boldsymbol{X} \sim \mathrm{Mult}_k(n, \boldsymbol{p})$, then the joint PMF of $\boldsymbol{X}$ is

$$P(X_1 = n_1, \ldots, X_k = n_k) = \frac{n!}{n_1! n_2! \ldots n_k!} \cdot p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k},$$

for $n_1, \ldots, n_k$ satisfying $n_1 + \cdots + n_k = n$.

Multinomial joint and marginal distributions

If $\boldsymbol{X} \sim \mathrm{Mult}_k(n, \boldsymbol{p})$, then the joint PMF of $\boldsymbol{X}$ is

$$P(X_1 = n_1, \ldots, X_k = n_k) = \frac{n!}{n_1! n_2! \ldots n_k!} \cdot p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k},$$

for $n_1, \ldots, n_k$ satisfying $n_1 + \cdots + n_k = n$.

The marginals of a Multinomial are Binomial. Specifically, if $\boldsymbol{X} \sim \mathrm{Mult}_k(n, \boldsymbol{p})$, then $X_j \sim \mathrm{Bin}(n, p_j)$.

Multinomial lumping and conditioning

If $\boldsymbol{X} \sim \mathrm{Mult}_k(n, \boldsymbol{p})$, then for any distinct i and j, $X_i + X_j \sim \mathrm{Bin}(n, p_i + p_j)$. The random vector of counts obtained from merging categories i and j is still Multinomial. For example, merging categories 1 and 2 gives

$$(X_1 + X_2, X_3, \ldots, X_k) \sim \mathrm{Mult}_{k-1}(n, (p_1 + p_2, p_3, \ldots, p_k)).$$

Multinomial conditioning gives another Multinomial. For example, given that there are n_1 objects in category 1, the remaining $n - n_1$ objects fall into categories 2 through k,

$$(X_2, \ldots, X_k) \mid X_1 = n_1 \sim \mathrm{Mult}_{k-1}(n - n_1, (p'_2, \ldots, p'_k)),$$

where $p'_j = p_j/(p_2 + \cdots + p_k)$, the conditional probability of an object falling into category j given that it is not in category 1.

Covariance in a Multinomial

Let $(X_1, \ldots, X_k) \sim \mathrm{Mult}_k(n, \boldsymbol{p})$, where $\boldsymbol{p} = (p_1, \ldots, p_k)$. For $i \neq j$, $\mathrm{Cov}(X_i, X_j) = -np_ip_j$.

Covariance in a Multinomial

Let $(X_1, \ldots, X_k) \sim \mathrm{Mult}_k(n, \boldsymbol{p})$, where $\boldsymbol{p} = (p_1, \ldots, p_k)$. For $i \neq j$, $\mathrm{Cov}(X_i, X_j) = -np_ip_j$.

Let $i = 1$ and $j = 2$. We know $X_1 + X_2 \sim \mathrm{Bin}(n, p_1 + p_2)$, $X_1 \sim \mathrm{Bin}(n, p_1)$, and $X_2 \sim \mathrm{Bin}(n, p_2)$. Therefore

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2\mathrm{Cov}(X_1, X_2)$$

becomes

$$n(p_1+p_2)(1-(p_1+p_2)) = np_1(1-p_1)+np_2(1-p_2)+2\mathrm{Cov}(X_1, X_2).$$

Solving for $\mathrm{Cov}(X_1, X_2)$ gives $\mathrm{Cov}(X_1, X_2) = -np_1p_2$. By the same logic, for $i \neq j$, we have $\mathrm{Cov}(X_i, X_j) = -np_ip_j$.

Correlation

The correlation between random variables X and Y is

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

This is undefined when $\mathrm{Var}(X) = 0$ or $\mathrm{Var}(Y) = 0$.

Shifting and scaling of X and Y have no effect on their correlation, since shifting does not affect $\mathrm{Cov}(X, Y)$, $\mathrm{Var}(X)$, or $\mathrm{Var}(Y)$, and

$$\mathrm{Corr}(cX, Y) = \frac{\mathrm{Cov}(cX, Y)}{\sqrt{\mathrm{Var}(cX)\mathrm{Var}(Y)}} = \frac{c\mathrm{Cov}(X, Y)}{\sqrt{c^2\mathrm{Var}(X)\mathrm{Var}(Y)}} = \mathrm{Corr}(X, Y).$$

Correlation bounds

For any random variables X and Y,

$$-1 \leq \mathrm{Corr}(X, Y) \leq 1.$$

Correlation bounds

For any random variables X and Y,

$$-1 \leq \mathrm{Corr}(X, Y) \leq 1.$$

Without loss of generality we can assume X and Y have variance 1, since scaling does not change the correlation.

Let $\rho = \mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y)$. Using the fact that variance is nonnegative, we have

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y) = 2 + 2\rho \geq 0, \\
\mathrm{Var}(X-Y) &= \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X, Y) = 2 - 2\rho \geq 0.
\end{aligned}$$

Thus, $-1 \leq \rho \leq 1$.

Example: Exponential max and min

Let X and Y be i.i.d. $\mathrm{Expo}(1)$ random variables. Find the correlation between $\max(X, Y)$ and $\min(X, Y)$.

Bivariate Normal distribution

Let $-\infty < \mu_X < \infty$, $-\infty < \mu_Y < \infty$, $\sigma_X > 0$, $\sigma_Y > 0$, and $-1 < \rho < 1$ be five real numbers. The **Bivariate Normal** PDF with means μ_X and μ_Y, variances σ_X^2 and σ_Y^2, and correlation ρ is the bivariate PDF given by

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

for $-\infty < x < \infty$, $-\infty < y < \infty$. The many nice properties of this distribution include:

1. The marginal distribution of X is $\mathcal{N}(\mu_X, \sigma_X^2)$.
2. The marginal distribution of Y is $\mathcal{N}(\mu_Y, \sigma_Y^2)$.
3. The correlation between X and Y is ρ.
4. For any constants a and b, the distribution of $aX + bY$ is $\mathcal{N}(a\mu_X + b\mu_Y, a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\rho\sigma_X\sigma_Y)$

Marginal distribution of X

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\
&= \frac{\exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2(1-\rho^2)}\right)}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(\frac{\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y(1-\rho^2)} - \frac{(y-\mu_Y)^2}{2\sigma_Y^2(1-\rho^2)}\right) dy \\
&= \frac{\exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right)}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{(y-\mu_Y-\frac{\rho\sigma_Y}{\sigma_X}(x-\mu_X))^2}{2\sigma_Y^2(1-\rho^2)}\right) dy \\
&= \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right)
\end{aligned}
$$

which is the PDF of $\mathcal{N}(\mu_X, \sigma_X^2)$.

Similarly, the marginal PDF of Y is that of $\mathcal{N}(\mu_Y, \sigma_Y^2)$.

Correlation between X and Y

By definition,

$$\begin{aligned}\mathrm{Corr}(X,Y) &= \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = E\left(\frac{X-\mu_X}{\sigma_X}\right)\left(\frac{Y-\mu_Y}{\sigma_Y}\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) f(x,y)dxdy.\end{aligned}$$

Letting $s = \left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)$ and $t = \frac{x-\mu_X}{\sigma_X}$, we obtain

$$\begin{aligned}\mathrm{Corr}(X,Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} sf\left(\sigma_X t + \mu_X, \sigma_Y \frac{s}{t} + \mu_Y\right)\frac{\sigma_X\sigma_Y}{|t|}dsdt \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{s\exp\left(-\frac{t^2-2\rho s+s^2/t^2}{2(1-\rho^2)}\right)}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\frac{\sigma_X\sigma_Y}{|t|}dsdt \\ &= \int_{-\infty}^{\infty}\frac{\exp(-t^2/2)}{2\pi\sqrt{(1-\rho^2)t^2}}\left[\int_{-\infty}^{\infty} s\exp\left(-\frac{(s-\rho t^2)^2}{2(1-\rho^2)t^2}\right)ds\right]dt \\ &= \frac{\rho}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t^2\exp(-t^2/2)dt = \rho.\end{aligned}$$