Math 40 Probability and Statistical Inference Winter 2021

Yoonsang Lee (yoonsang.lee@dartmouth.edu)

Lecture 10: Bivariate Normal (3.5)

・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・

Bivariate normal distribution (3.5)

Now our random variable X has two components, X_1 and X_2 . X is called a bivariate random variable.

Let X_1 and X_2 have mean μ_1 and μ_2 and variances σ_1^2 and σ_2^2 respectively. We further assume that the correlation coefficient between X_1 and X_2 is ρ .

The bivariate normal distribution of $X = (X_1, X_2)$ is given by

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

・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・

- If X₁ and X₂ are uncorrelated, then X₁ and X₂ are independente (this fact holds for only bivariate normal distributions)
- Principal axis and minor principal axis.

Bivariate normal distribution (3.5)

Example 3.54 Oil spill. An oil spill happened in the ocean. Gusty wind and rough sea moves the oil spill toward the shore. The distance covered in the x and y directions (meters per hour) follows an exponential distribution with $\lambda = 1/150$ and $\lambda = 1/100$ with probability 3/5 and 1/2, respectively. Approximate the probability that the oil spill does not reach the shoreline given by the equation x + y = 3000 using the CLT.

・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・

Regression as conditional mean (3.5.1)

- ► Let Y and X follows a bivariate normal distribution with mean (μ_y, μ_x) and covariance matrix $\begin{pmatrix} \sigma_y^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_z^2 \end{pmatrix}$.
- We want to calculate the conditional expectation of Y given X.

►
$$Y|(X = x) \sim \mathcal{N}(\mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x), \sigma_{y|x}^2)$$
 where
 $\sigma_{y|x}^2 = \sigma_y^2(1 - \rho^2)$

Thus, the conditional expectation is

$$E(Y|X=x) = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x-\mu_x)$$

which is the least squares linear regression.

Variance decomposition and coefficient of determination (3.5.2)

We have seen the variance decomposition

$$Var(Y) = E(Var(Y|X)) + Var(E(Y|X)).$$

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ 三 のへぐ

In the case of a bivariate normal distribution,

►
$$Var(Y) = \sigma_y^2$$

► $E(Var(Y|X)) = \sigma_{y|x}^2 = \sigma_y^2(1 - \rho^2)$
► $Var(E(Y|X)) = Var(\mu_y + \rho \frac{\sigma_y}{\sigma_x}(X - \mu_x)) = \sigma_y^2 \rho^2$

Copula (3.5.4)

Goal: we know marginal distributions F_X and F_Y . From this information, we want to estimate the joint density of X and Y.

- Let $\phi(u, v; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(u^2 2\rho uv + v^2)}$, the standard bivariate normal with the correlation ρ .
- The copula cdf is given by

$$F(x,y;\rho) = \int_{-\infty}^{\Phi^{-1}(F_X(x))} \int_{-\infty}^{\Phi^{-1}(F_Y(y))} \phi(u,v;\rho) du dv$$

where Φ is the univariate standard normal cdf.

What about the coupla density? Take the derivatives with respect to x and y, which yields

$$f(x, y; \rho) = \frac{\phi(\Phi^{-1}(F_X(x)), \Phi^{-1}(F_Y(y)); \rho)}{\phi(\Phi^{-1}(F_X(x)))\phi(\Phi^{-1}(F_Y(y)))} f_X(x) f_Y(y)$$

Inverse function theorem: $(f^{-1})' = \frac{1}{f'(f^{-1}(b))}$.