# Math 40 Probability and Statistical Inference 

 Winter 2021Yoonsang 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 $\mu_{1}$ and $\mu_{2}$ and variances $\sigma_{1}^{2}$ and $\sigma_{2}^{2}$ respectively. We further assume that the correlation coefficient between $X_{1}$ and $X_{2}$ is $\rho$.
The bivariate normal distribution of $X=\left(X_{1}, X_{2}\right)$ is given by

$$
f\left(x_{1}, x_{2}\right)=\frac{1}{2 \pi \sigma_{1} \sigma_{2} \sqrt{1-\rho^{2}}} e^{-\frac{1}{1-\rho^{2}}\left[\left(\frac{x_{1}-\mu_{1}}{\sigma_{1}}\right)^{2}-2 \rho\left(\frac{x_{1}-\mu_{1}}{\sigma_{1}}\right)\left(\frac{x_{2}-\mu_{2}}{\sigma_{2}}\right)+\left(\frac{x_{2}-\mu_{2}}{\sigma_{2}}\right)^{2}\right]}
$$

- If $X_{1}$ and $X_{2}$ are uncorrelated, then $X_{1}$ and $X_{2}$ 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 $\left(\mu_{y}, \mu_{x}\right)$ and covariance matrix $\left(\begin{array}{cc}\sigma_{y}^{2} & \rho \sigma_{x} \sigma_{y} \\ \rho \sigma_{x} \sigma_{y} & \sigma_{x}^{2}\end{array}\right)$.
- We want to calculate the conditional expectation of $Y$ given $X$.
- $Y \left\lvert\,(X=x) \sim \mathcal{N}\left(\mu_{y}+\rho \frac{\sigma_{y}}{\sigma_{x}}\left(x-\mu_{x}\right), \sigma_{y \mid x}^{2}\right)\right.$ where

$$
\sigma_{y \mid x}^{2}=\sigma_{y}^{2}\left(1-\rho^{2}\right)
$$

- Thus, the conditional expectation is

$$
E(Y \mid X=x)=\mu_{y}+\rho \frac{\sigma_{y}}{\sigma_{x}}\left(x-\mu_{x}\right)
$$

which is the least squares linear regression.

Variance decomposition and coefficient of determination (3.5.2)

We have seen the variance decomposition

$$
\operatorname{Var}(Y)=E(\operatorname{Var}(Y \mid X))+\operatorname{Var}(E(Y \mid X))
$$

In the case of a bivariate normal distribution,

- $\operatorname{Var}(Y)=\sigma_{y}^{2}$
- $E(\operatorname{Var}(Y \mid X))=\sigma_{y \mid x}^{2}=\sigma_{y}^{2}\left(1-\rho^{2}\right)$
- $\operatorname{Var}(E(Y \mid X))=\operatorname{Var}\left(\mu_{y}+\rho \frac{\sigma_{y}}{\sigma_{x}}\left(X-\mu_{x}\right)\right)=\sigma_{y}^{2} \rho^{2}$


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

$$
F(x, y ; \rho)=\int_{-\infty}^{\Phi^{-1}\left(F_{X}(x)\right)} \int_{-\infty}^{\Phi^{-1}\left(F_{Y}(y)\right)} \phi(u, v ; \rho) d u d v
$$

where $\Phi$ is the univariate standard normal cdf.

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

$$
f(x, y ; \rho)=\frac{\phi\left(\Phi^{-1}\left(F_{X}(x)\right), \Phi^{-1}\left(F_{Y}(y)\right) ; \rho\right)}{\phi\left(\Phi^{-1}\left(F_{X}(x)\right)\right) \phi\left(\Phi^{-1}\left(F_{Y}(y)\right)\right)} f_{X}(x) f_{Y}(y)
$$

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

