Week 11: Bivariate Gaussian random variables, multivariate random variables, random samples, Central Limit Theorem

Armenak Petrosyan

Tuesday class

- Let (X, Y) be a bivariate random variable.
- Note the relationship between conditional and total means:

$$\mu_Y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x,y) \, \mathrm{d}y \, \mathrm{d}x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y h(y|x) \, f_X(x) \, \mathrm{d}y \, \mathrm{d}x$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y h(y|x) \, \mathrm{d}y \, f_X(x) \, \mathrm{d}x$$

$$= \int_{-\infty}^{\infty} E[Y|x] \, f_X(x) \, \mathrm{d}x.$$

Law of total variance

Theorem

$$\sigma_Y^2 = \int\limits_{-\infty}^{\infty} \sigma_{Y|x}^2 \, f_X(x)\, \mathrm{d}x + \int\limits_{-\infty}^{\infty} (E[Y|x] - \mu_Y)^2 \, f_X(x)\, \mathrm{d}x.$$

Proof

$$\begin{aligned}
\int\limits_{-\infty}^{\infty} \sigma_{Y|x}^2 \, f_X(x)\, \mathrm{d}x &= \int\limits_{-\infty}^{\infty} \left[\int\limits_{-\infty}^{\infty} E[Y^2|x] - E[Y|x]^2 \right] f_X(x)\, \mathrm{d}x \\
&= \int\limits_{-\infty}^{\infty} \left[\int\limits_{-\infty}^{\infty} y^2 \, h(y|x)\, \mathrm{d}y - E[Y|x]^2 \right] f_X(x)\, \mathrm{d}x \\
&= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} y^2 \, h(y|x)\, f_X(x)\, \mathrm{d}y\, \mathrm{d}x - \int\limits_{-\infty}^{\infty} E[Y|x]^2 \, f_X(x)\, \mathrm{d}x \\
&= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} y^2 \, f(x,y)\, \mathrm{d}y\, \mathrm{d}x - \int\limits_{-\infty}^{\infty} E[Y|x]^2 \, f_X(x)\, \mathrm{d}x
\end{aligned}$$

Proof (cont.)

$$\int_{-\infty}^{\infty} (E[Y|x] - \mu_Y)^2 \, f_X(x) \, \mathrm{d}x$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} E[Y|x]^2 \, f_X(x) \, \mathrm{d}x - \int_{-\infty}^{\infty} 2\mu_Y \, E[Y|x] \, f_X(x) \, \mathrm{d}x + \int_{-\infty}^{\infty} \mu_Y^2 \, f_X(x) \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} E[Y|x]^2 \, f_X(x) \, \mathrm{d}x - 2\mu_Y \int_{-\infty}^{\infty} E[Y|x] \, f_X(x) \, \mathrm{d}x + \mu_Y^2 \int_{-\infty}^{\infty} f_X(x) \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} E[Y|x]^2 \, f_X(x) \, \mathrm{d}x - 2\mu_Y \cdot \mu_Y + \mu_Y^2 \cdot 1 \\
&= \int_{-\infty}^{\infty} E[Y|x]^2 \, f_X(x) \, \mathrm{d}x - \mu_Y^2 .
\end{aligned}$$

Proof (cont.)

Finally,

$$\begin{aligned}
&\int_{-\infty}^{\infty} \sigma^2_{Y|x}\, f_X(x)\,\mathrm{d}x + \int_{-\infty}^{\infty} (E[Y|x]-\mu_Y)^2\, f_X(x)\,\mathrm{d}x \\
=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2\, f(x,y)\,\mathrm{d}y\,\mathrm{d}x - \int_{-\infty}^{\infty} E[Y|x]^2\, f_X(x)\,\mathrm{d}x + \int_{-\infty}^{\infty} E[Y|x]^2\, f_X(x)\,\mathrm{d}x - \mu_Y^2 \\
=&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2\, f(x,y)\,\mathrm{d}y\,\mathrm{d}x - \mu_Y^2 = E[Y^2]-\mu_Y^2 \\
=&\mathrm{Var}(Y).
\end{aligned}$$

$\square$

Bivariate normal distributions: motivation

Consider the following model: let (X, Y) be bivariate continuous random variable and

1) X is normally distributed:
2) For every $x \in \mathbb{R}$, the conditional pdf of Y given x is a normal distribution.
3) $E[Y|x]$ is a linear function of x.
4) $\sigma_{Y|x} = \sigma$ is constant (does not depend on x).

- Condition 1) implies

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}.$$

- Condition 2) and 4) imply

$$h(y|x) = \frac{1}{\sigma_{Y|x}\sqrt{2\pi}} e^{-\frac{(y-E[Y|x])^2}{2\sigma_{Y|x}^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-E[Y|x])^2}{2\sigma^2}}.$$

- Condition 3) implies

$$E[Y|x] = \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X) + \mu_Y.$$

- Condition 3) and 4) imply (using the law of total variance)

$$\sigma_Y^2 = \int_{-\infty}^{\infty} \sigma_{Y|x}^2 \, f_X(x) \, \mathrm{d}x + \int_{-\infty}^{\infty} (E[Y|x] - \mu_Y)^2 \, f_X(x) \, \mathrm{d}x$$

$$= \int_{-\infty}^{\infty} \sigma^2 \, f_X(x) \, \mathrm{d}x + \int_{-\infty}^{\infty} \left(\rho \frac{\sigma_Y}{\sigma_X} (x - \mu_X) \right)^2 f_X(x) \, \mathrm{d}x$$

$$= \sigma^2 + \left(\rho \frac{\sigma_Y}{\sigma_X} \right)^2 \int_{-\infty}^{\infty} (x - \mu_X)^2 \, f_X(x) \, \mathrm{d}x = \sigma^2 + \rho^2 \sigma_Y^2 .$$

$$\boxed{\sigma^2 = \sigma_Y^2 (1 - \rho^2)}.$$

Wrap-up

Thus, we have

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} = \frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}$$

$$\begin{aligned} h(y|x) &= \frac{1}{\sigma_Y\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{\left(y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right)^2}{2\sigma_Y^2(1-\rho^2)}} \\ &= \frac{1}{\sigma_Y\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_Y}{\sigma_Y}-\rho\frac{x-\mu_X}{\sigma_X}\right)^2} \end{aligned}$$

Therefore

$$\begin{aligned} f(x,y) &= h(y|x)f_X(x) \\ &= \frac{1}{2\pi\,\sigma_Y\sigma_X\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_Y}{\sigma_Y}-\rho\frac{x-\mu_X}{\sigma_X}\right)^2-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \\ &= \frac{1}{2\pi\,\sigma_Y\sigma_X\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2-2\rho\frac{y-\mu_Y}{\sigma_Y}\frac{x-\mu_X}{\sigma_X}+\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]}. \end{aligned}$$

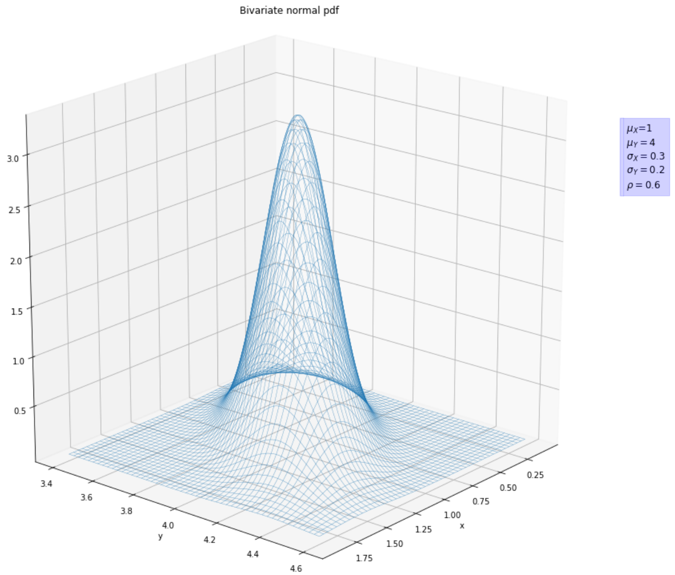
Bivariate normal pdf (definition)

Definition

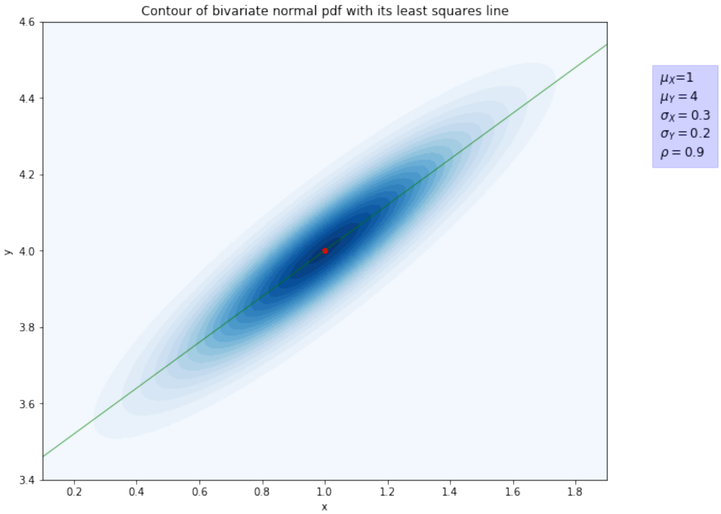
The joint pdf

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

is called **bivariate normal pdf**.



Horizontal, vertical (in fact all) slices are Gaussians.



The spread around the least squares line (in conditional probability)

$$\boxed{\sigma = \sigma_Y(1-\rho^2)}$$

For bivariate normal pdf, we saw

$$f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}.$$

By symmetry (interchanging X and Y),

$$f_Y(y) = \frac{1}{\sigma_Y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2}.$$

From the formulas of joint and marginal pdf-s, we have

Theorem

For bivariate normal distribution, X and Y are independent if and only if they are not correlated: $\rho = 0$.

Standard bivariate normal pdf

Definition

If $\mu_X, \mu_Y, \rho = 0$ and $\sigma_X = \sigma_Y = 1$ in the bivariate normal pdf, i.e.

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

then it is called **bivariate standard normal pdf.**

- In bivariate standard normal distribution, X and Y are independent and have identical $N(0, 1)$ distributions.

- Let $X_1, \dots, X_n$ be random variables defined on the same set of outcomes S.

Definition

Assume all X_i are discrete. Then

$$f(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$$

is called their **joint pmf**.

Definition

The multivariate random variable $(X_1, \dots, X_n)$ is called **continuous** if there exists a function $f : \mathbb{R}^n \to \mathbb{R}$ such that

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(z_1, \dots, z_n) \, \mathrm{d}z_1 \dots \mathrm{d}z_n.$$

$f(z_1, \dots, z_n)$ is called the **joint pdf** of $X_1, \dots, X_n$.

Marginal pmf and pdf

Define the marginal pmf for discrete case as

$$f_{X_i}(x) = P(X_i = x)$$

and the marginal pdf for continuous as

$$f_{X_i}(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1}\, \mathrm{d}x\, \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n.$$

Thursday class

Definition

$X_1, \dots, X_n$ are called independent, if for any $(x_1, \dots, x_n)$

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

where f-s are the pdf-s in the continuous case and the pmf-s in the discrete case.

- If $X_1, \dots, X_n$ are independent then for any functions $u_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \dots, n$, the random variables $Y_i = u_i(X_i)$ are also independent.

Theorem

If $X_1, \dots, X_n$ are independent then any subset of them are also independent.

- In particular, any pair (X_i, X_j) for $i \neq j$ are independent.
- Therefore,

$$\mathrm{Cov}(X_i, X_j) = \begin{cases} \mathrm{Var}(X_i) & i = j \\ 0 & i \neq j \end{cases}.$$

Expectation

For any function $u : \mathbb{R}^n \to \mathbb{R}$,

$$E[u(X_1, \ldots, X_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u(x_1, \ldots, x_n) f(x_1, \ldots, x_n) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n \quad \text{(for continuous)}$$

$$E[u(X_1, \ldots, X_n)] = \sum_{x_1 \in \mathrm{Range}(X_1)} \cdots \sum_{x_n \in \mathrm{Range}(X_n)} u(x_1, \ldots, x_n) f(x_1, \ldots, x_n) \quad \text{(for discrete)}$$

Theorem

If $X_1, \ldots, X_n$ are independent then, for any functions $u_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, n$,

$$E[u_1(X_1) \cdots u_n(X_n)] = E[u_1(X_1)] \cdots E[u_n(X_n)].$$

Proof.

$$\begin{aligned}
E[u_1(X_1) \cdots u_n(X_n)] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u_1(X_1) \cdots u_n(X_n)\, f(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} u_1(x_1) \cdots u_n(x_n)\, f_1(x_1) \cdots f_n(x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n \\
&= \int_{-\infty}^{\infty} u_1(x_1) f_1(x_1)\, \mathrm{d}x_1 \cdots \int_{-\infty}^{\infty} u_n(x_n) f_n(x_n)\, \mathrm{d}x_n \\
&= E[u_1(X_1)] \cdots E[u_n(X_n)].
\end{aligned}$$

□

Theorem

Suppose $X_1, \dots, X_n$ are independent and $Y = \alpha_1 X_1 + \cdots + \alpha_n X_n$. Then

$$E[Y] = \sum_{i=1}^{n} \alpha_i \mu_i$$

$$\sigma_Y^2 = \sum_{i=1}^{n} \alpha_i^2 \sigma_{X_i}^2.$$

Proof

$$\begin{aligned}
E[\sum_{i=1}^{n} \alpha_i X_i] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} [\sum_{i=1}^{n} \alpha_i x_i]\, f(x_1, \dots, x_n)\, \mathrm{d}x_1 \dots \mathrm{d}x_n \\
&= \sum_{i=1}^{n} \alpha_i \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i\, f(x_1, \dots, x_n)\, \mathrm{d}x_1 \dots \mathrm{d}x_n \\
&= \sum_{i=1}^{n} \alpha_i \mu_i
\end{aligned}$$

Proof (cont.)

$$
\begin{aligned}
\sigma_Y^2 &= E[Y^2 - E[Y]^2] = E[(\sum_{i=1}^n \alpha_i X_i - \sum_{i=1}^n \alpha_i \mu_i)^2] \\
&= E[(\sum_{i=1}^n \alpha_i (X_i - \mu_i))^2] \\
&= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j E[(X_i - \mu_i)(X_j - \mu_j)] \\
&= \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \operatorname{Cov}(X_i, X_j) \\
&= \sum_{i=1}^n \alpha_i^2 \sigma_{X_i}^2 .
\end{aligned}
$$

□

- If two random variables have the same distribution then they have the same mean, variance, moments etc, because all these quantities are computed using the pmf or pdf only.

Definition

If $X_1, \ldots, X_n$ are independent and have the same (marginal) distributions,

$$f_{X_1}(x) = \cdots = f_{X_n}(x),$$

then they are called **independent identically distributed** or **i.i.d.** random variables.

- For i.i.d. random variables, we say $X_1, \ldots, X_n$ form a **random sample of size** n from the common distribution.
- We use the values $X_1(s), \ldots, X_n(s)$ to model random samples taken during a single run of the experiment.

Example

The laboratory assistant catches (randomly samples) n insects of the same type during the experiment and measures their wing lengths. The corresponding lengths will be $X_1(s), \ldots, X_n(s)$.
If another assistant at a different laboratory does the same experiment, his measurement may be $X_1(s'), \ldots, X_n(s')$ for potentially different from s value of s'.

Empirical mean of a random sample

Definition

Let $X_1, \dots, X_n$ be a random sample (i.i.d. random variables) with mean μ and variance σ. The random variable

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

is called the **empirical mean** of the sample or **sample mean**.

- In this case μ is called **population mean**.
- When n is fixed, we prefer $\bar{X}$ instead of $\bar{X}_n$.
- Sample mean is a random variable.

Notice that

$$E[\bar{X}_n] = E[\frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n] = \frac{1}{n}n\mu = \mu.$$

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}(\frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n) = \frac{1}{n^2}n\sigma^2 = \frac{1}{n}\sigma^2.$$

- As n increases, the mean stays the same and the variance decreases.
- Therefore, we expect $\bar{X}_n$ to accumulate around the mean.

Theorem (Strong law of large numbers)

Let $X_1, \ldots, X_n$ be a random sample (i.i.d. random variables) with mean μ and variance σ then

$$P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1.$$

- In other words, the sample mean converges to the population mean with probability equal to 1.

- A distribution is uniquely determined by its moment generating function.
- To understand how $\bar{X}_n$ is distributed, we standardize it by taking its Z-score:

$$\bar{Z}_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} X_i - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} (X_i - \mu)}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}}{n}\sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i.$$

- And we estimate the distribution of $\bar{X}_n$ by looking at the mgf of $\bar{Z}_n$.

$$M_{\bar{Z}_n}(t) = E[e^{t\bar{Z}_n}] = E\left[e^{t\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n} Z_i}\right] = E\left[\prod_{i=1}^{n} e^{\frac{t}{\sqrt{n}}Z_i}\right]$$

$$\overset{\text{(independ.)}}{=} \prod_{i=1}^{n} E\left[e^{\frac{t}{\sqrt{n}}Z_i}\right] = \left(M\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

where $M(t)$ is the mgf of Z_i which is same for all Z_i since they are identically distributed.

Fact from Calculus (Taylor's theorem):

If function $M(t)$ has up to second derivative at 0, then

$$M(t) = M(0) + \frac{M'(0)}{1!}t + \frac{M''(0)}{2!}t^2 + h(t) \cdot t^2$$

where $h(t) \to 0$ when $t \to 0$.

- Note that
$$M(0) = E[e^0] = 1, \quad M'(0) = E[Z_i] = 0, \quad M''(0) = E[Z_i^2] = \mathrm{Var}(Z_i) + E[Z_i]^2 = 1.$$
- Therefore
$$M(t) = 1 + \frac{1}{2!}t^2 + h(t) \cdot t^2.$$

Fact from Calculus:

If $a_n \to a$ then
$$\left(1 + \frac{a_n}{n}\right)^n \to e^a.$$

- Therefore, for $a_n = \frac{t^2}{2} + h\left(\frac{t}{\sqrt{n}}\right)t^2 \to \frac{t^2}{2}$ as $n \to \infty$,
$$M_{\bar{Z}_n}(t) = \left(1 + \frac{t^2}{2n} + h\left(\frac{t}{\sqrt{n}}\right) \cdot \frac{t^2}{n}\right)^n \to e^{\frac{t^2}{2}}$$

- When n is large, $M_{\bar{Z}_n}(t)$ is close to $e^{\frac{t^2}{2}}$.
- $e^{\frac{t^2}{2}}$ is the mgf of standard normal distribution.
- The distribution of a random variable is uniquely determined by its mgf.

In conclusion:

When n is large, $\bar{Z}_n$'s distribution is close to $N(0,1)$, or $\bar{X}_n$'s distribution is close to $N(\mu, \frac{\sigma^2}{n})$.