Lecture 10 - Continuous Random Variables II

Joint distribution of a continuous random vector

- Random vector $\mathbf{X} : \Omega \to \mathbb{R}^n$ is called continuous, if there exists a function $f : \mathbb{R}^n \to \mathbb{R}_+$ such that for any $A \subseteq \mathbb{R}^n$:

$$P(\mathbf{X} \in A) = \int_A f(\mathbf{x}) d\mathbf{x}$$

- Function f is called the probability density function of the vector $\mathbf{X}$.
- Remark: the above integral is multidimensional
- In the case of a two element vector $\mathbf{X} = (X, Y)$ we have density $f(x, y)$, and the normalization:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$$

Marginal and conditional density for two variables

- Given a joint density $f(x, y)$ of variables X and Y we define
  - Marginal density:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

  - Conditional density, e.g.:

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$$

- Conditional expectation, e.g.:

$$E(Y|X=x) = \int_{\infty}^{\infty} y f_{Y|X}(y|x) dx$$

- The above definitions generalize easily two more than two variables

The law of total probability

- Given a joint density $f(x,y)$ of variables X and Y:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)dx \\ &= \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x) dx \end{aligned}$$

Independent continuous random variables

- Continuous random variables X and Y are independent if and only if:

$$f(x,y) = f_X(x) f_Y(y)$$

- Or, in general:

$$f(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$$

General rules for any random variables

- The following generalizes to continuous random variables:
- For **any** random variables:

$$E(X+Y) = EX + EY$$
$$D^2(X \pm Y) = D^2(X) \pm 2C(X,Y) + D^2(Y)$$

- For independent random variables

$$E(XY) = (EX)(EY)$$
$$C(X, Y) = 0$$
$$D^2(X \pm Y) = D^2(X) + D^2(Y)$$

- We can write similar formulas for more than two random variables

Maximum and minimum of independent random variables

- $X_1, X_2, \ldots, X_n$ — independent random variables with the same probability distribution with cumulative distribution function F_X.
- Define $Y = \max\{X_1, \ldots, X_n\}$ and $Z = \min\{X_1, \ldots, X_n\}$
- The c.d.f. F_Y and F_Z can be calculated as follows:

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(\max\{X_1, \ldots, X_n\} \le y) \\ &= P(X_1 \le y, X_2 \le y, \ldots, X_n \le y) \\ &=^{(*)} P(X_1 \le y) \cdot \ldots \cdot P(X_n \le y) \\ &= F_X(y)^n \end{aligned}$$

- Where (*) means we used the independence of $X_1, \ldots, X_n$.
- Similarly we get: $F_Z(z) = 1 - (1 - F_X(z))^n$
- In case when $X_1, \ldots, X_n$ are random variables distributed with uniform distribution $\mathrm{Unif}[0, 1]$, then we can obtain the distributions of Y and Z by:

$$f_Y(y) = ny^{n-1}$$
$$f_Z(z) = n(1 - z)^{n-1}$$

The distribution of a sum of independent random variables

- X, Y — independent, continuous random variables with densities f_X and f_Y. We calculate the density of $Z = X + Y$ by:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx$$

Convolution

- A convolution of functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ is a function:

$$(f * g)(x) = \int_{-\infty}^{\infty} f(t) g(x - t) dt$$

- We can see that the density of $Z = X + Y$ (where X and Y are independent) is a convolution of densities f_X and f_Y:

$$f_Z(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(t) f_Y(z - t) dt$$

Normal random variables

- Consider independent variables $X_1, \ldots, X_n$ where:

$$X_i \sim N(\mu_i, \sigma_i^2), \quad i = 1, \ldots, n$$

- We define $Z = \sum_{i=1}^{n} a_i X_i$ for any numbers $a_1, \ldots, a_n$. Then this random variable has a distribution:

$$Z \sim N\Big(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\Big)$$

- Conclusion: any linear combination of independent normal random variables is a normal random variable. In a special case, the sum of two normal random variables is a normal random variable.

Chi-square distribution

- Variable Z has a “chi-square” distribution with k degrees of freedom, if Z can be written as a sum of squared of k independent random variables with distribution $N(0,1)$:

$$Z = \sum_{i=1}^{k} X_i^2, \quad X_i \sim N(0,1), \text{ independent}$$

- We denote $Z \sim \chi^2(k)$.
- The expected value of Z is then $EZ = k$.

t-Student distribution

- A random variable T has a t-Student distribution with k degrees of freedom, if T can be written as:

$$T = \frac{X}{\sqrt{Z}}\sqrt{k}$$

- Where
  - $X \sim N(0,1)$
  - $Z \sim \chi^2(k)$
  - X and Z are independent
- We denote $T \sim t(k)$.
- For $k \to \infty$ the t-Student distribution converges to normal distribution $N(0,1)$.