MULTIVARIATE STATISTICS, Lesson 3.
C.d.f. of the multivariate normal distribution

Numerical approximations for

$$F(x_1, \dots, x_p) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_p} f(t_1, \dots, t_p)\, dt_1, \dots dt_p.$$

1. *Monte Carlo method.* Approximate the probability

$$F(x_1, \dots, x_p) = \mathbb{P}(X_1 < x_1, \dots, X_p < x_p)$$

with the corresponding relative frequency based on an $\mathbf{X}_1, \dots, \mathbf{X}_n \sim \mathcal{N}_p(\mathbf{m}, \mathbf{C})$ i.i.d. sample. How to do it?

2. *Expansion by means of Hermite polynomials*

- **Definition (correlation matrix):** $\mathbf{R} = \mathbf{D}^{-1/2}\mathbf{C}\mathbf{D}^{-1/2}$, where the diagonal matrix $\mathbf{D}$ contains the positive diagonal entries of the covariance matrix $\mathbf{C}$ in its main diagonal.
- **Definition** (l^{th} orthogonal Hermite polynomial):

$$H_l(x) = (-1)^l e^{x^2/2} \frac{d^l}{dx^l} e^{-x^2/2}, \quad (l = 0, 1, 2, \dots)$$

- **Proclaim:** Let $\mathbf{X} = (X_1, \dots, X_p) \sim \mathcal{N}_p(\mathbf{0}, \mathbf{R})$, and suppose that the eigenvalues of the matrix $\mathbf{R} - \mathbf{I}$ are less then 1 in absolute value (or equivalently, the spectrum of the correlation matrix $\mathbf{R}$ is in the (0,2) interval). Then

$$\mathbb{P}(X_1 \geq x_1, \dots, X_p \geq x_p) = \prod_{m=1}^{p} (1 - \Phi(x_m)) +$$

$$+ \prod_{m=1}^{p} \phi(x_m) \cdot \sum_{k=1}^{\infty} \sum_{k_{ij}} \left(\prod_{i=1}^{p-1} \prod_{j=i+1}^{p} \frac{r_{ij}^{k_{ij}}}{k_{ij}!} \right) \prod_{q=1}^{p} H_{l_q - 1}(x_q),$$

where the summation if for k_{ij}'s such that

$$k_{ij} \geq 0 \text{ integer}, \quad \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} k_{ij} = k,$$

r_{ij}'s are entries of $\mathbf{R}$, further $l_q = \sum_{i=1}^{q-1} k_{iq} + \sum_{j=q+1}^{p} k_{qj}$, H_l is the l^{th} Hermite polynomial, and $H_{-1}(x) := 1 - \Phi(x)$. This series is absolutely and uniformly convergent over $\mathbb{R}^p$.