1.017/1.010 Class 11
Multivariate Probability

Multiple Random Variables

Recall the dart tossing experiment from Class 4. Treat the 2 dart coordinates as two different scalar random variables $\boldsymbol{x}$ and $\boldsymbol{y}$.

In this experiment the experimental outcome is the location where the dart lands. The random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ both depend on this outcome (they are defined over the same sample space). In this case we can define the following events:

$$A = [\boldsymbol{x}(\xi) \leq x] \quad B = [\boldsymbol{y}(\xi) \leq y] \quad C = [\boldsymbol{x}(\xi) \leq x,\ \boldsymbol{y}(\xi) \leq y] = A \cap B = AB$$

$\boldsymbol{x}$ and $\boldsymbol{y}$ are **independent** if A and B are independent events for all x and y:

$$P(C) = P(AB) = P(A)P(B)$$

Another example …

Consider a **time series** constructing from a sequence of random variables defined at different times (a series of *n* seismic observations or stream flows $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$, …, $\boldsymbol{x}_n$.). Each possible time series can be viewed as an outcome ξ of an underlying experiment. Events can be defined as above:

$$A_i = [\boldsymbol{x}_i(\xi) \leq x_i] \quad A_{ij} = [\boldsymbol{x}_i(\xi) \leq x_i,\ \boldsymbol{x}_j(\xi) \leq x_j] = A_i \cap A_j = A_i A_j$$

$\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are **independent** if:

$$P(A_{ij}) = P(A_i A_j) = P(A_i)P(A_j)$$

Multivariate Probability Distributions

Multivariate **cumulative distribution function** (CDF), for $\boldsymbol{x}, \boldsymbol{y}$ **continuous or discrete:**

$$F_{\boldsymbol{xy}}(x, y) = P[(\boldsymbol{x}(\xi) \leq x)(\boldsymbol{y}(\xi) \leq y)]$$

Multivariate **probability mass function** (PMF), for $\boldsymbol{x}, \boldsymbol{y}$ **discrete**:

$$p_{\boldsymbol{xy}}(x_i, y_j) = P[(\boldsymbol{x}(\xi) = x_i)(\boldsymbol{y}(\xi) = y_j)]$$

Multivariate **probability density function** (PDF), for $\boldsymbol{x}, \boldsymbol{y}$ **continuous**:

$$f_{xy}(x,y) = \frac{\partial^2 F_{xy}(x,y)}{\partial x \partial y}$$

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are **independent**:

$$F_{\boldsymbol{xy}}(x,y) = P[\boldsymbol{x} \le x]P[\boldsymbol{y} \le y] = F_{\boldsymbol{x}}(x)F_{\boldsymbol{y}}(y)$$
$$p_{\boldsymbol{xy}}(x_i, y_j) = p_{\boldsymbol{x}}(x_i)p_{\boldsymbol{y}}(y_j)$$
$$f_{\boldsymbol{xy}}(x,y) = f_{\boldsymbol{x}}(x)f_{\boldsymbol{y}}(y)$$

Computing Probabilities from Multivariate Density Functions

Probability that $(x, y) \in$ the region D:

$$P[(x,y) \in D] = \int_{(x,y) \in D} f_{xy}(x,y)\, dxdy$$

Covariance and Correlation

Dependence between random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ is frequently described with the **covariance** and **correlation**:

$$Cov(\boldsymbol{x},\boldsymbol{y}) = E[(\boldsymbol{x} - \overline{\boldsymbol{x}})(\boldsymbol{y} - \overline{\boldsymbol{y}})] = \int_{-\infty}^{+\infty} (\boldsymbol{x} - \overline{\boldsymbol{x}})(\boldsymbol{y} - \overline{\boldsymbol{y}}) f_{xy}(x,y)\, dx\, dy$$

$$Correl(\boldsymbol{x},\boldsymbol{y}) = \frac{Cov\ (\boldsymbol{x},\boldsymbol{y})}{[Var(\boldsymbol{x})Var(\boldsymbol{y})]^{1/2}} = \frac{Cov\ (\boldsymbol{x},\boldsymbol{y})}{Std(\boldsymbol{x})Std(\boldsymbol{y})}$$

Uncorrelated $\boldsymbol{x}$ and $\boldsymbol{y}$**:** $Cov(\boldsymbol{x},\boldsymbol{y}) = Correl\ (\boldsymbol{x},\boldsymbol{y}) = 0$

Independence implies uncorrelated (but not necessarily vice versa)

Examples

Two independent exponential random variables (parameters a_x and a_y):

$$f_{xy}(x,y) = f_x(x)f_y(y) = \frac{1}{a_x}\exp\left[-\frac{x}{a_x}\right]\frac{1}{a_y}\exp\left[-\frac{y}{a_y}\right] = \frac{1}{a_x a_y}\exp\left[-\frac{x}{a_x} - \frac{y}{a_y}\right]$$

$a_x = E(\boldsymbol{x})$, $a_y = E(\boldsymbol{y})$, $Correl(x,y) = 0$

Two dependent normally distributed random variables (parameters μ_x, μ_y, σ_x, σ_y, and ρ):

$$f_{xy}(x,y) = \frac{1}{2\pi |C|^{0.5}} \exp\left\{ -\left[\frac{(Z-\mu)' C^{-1} (Z-\mu)}{2} \right] \right\}$$

Z = vector of **random variables** = $[x \;\; y]'$

μ = vector of **means** = $[\, E(\boldsymbol{x}) \;\; E(\boldsymbol{y}) \,]'$

C = **covariance matrix** = $C = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$

$\sigma_x = Std(\boldsymbol{x}), \quad \sigma_y = Std(\boldsymbol{y}), \quad \rho = Correl(x,y)$

$|C|$ = **determinant** of $C = \sigma_x^2 \sigma_y^2 (1-\rho^2)$

C^{-1} = **inverse** of $C = \dfrac{1}{|C|} \begin{bmatrix} \sigma_y^2 & -\rho\sigma_x\sigma_y \\ -\rho\sigma_x\sigma_y & \sigma_x^2 \end{bmatrix}$

Multivariate probability distributions are rarely used except when:

1. The random variables are **independent**
2. The random variables are dependent but **normally distributed**

Exercise:

Use the MATLAB function `mvnrnd` to generate scatterplots of correlated bivariate normal samples. This function takes as arguments the means of x and y and the covariance matrix defined above (called `SIGMA` in the MATLAB documentation).

Assume $E[x] = 0$, $E[y] = 0$, $\sigma_x = 1$, $\sigma_y = 0$. Use `mvnrnd` to generate 100 (x, y) realizations . Use `plot` to plot each of these as a point on the (x,y) plane (do not connect the points). Vary the correlation coefficient ρ to examine its effect on the scatter. Consider $\rho = 0., 0.5, 0.9$. Use `subplot` to put plots for all 3 ρ values on one page.