Multiple Random Variables

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

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

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

$x$ and $y$ 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 $x_1, x_2, x_3, \ldots, x_n$). Each possible time series can be viewed as an outcome $\xi$ of an underlying experiment. Events can be defined as above:

$$A_i = [x_i(\xi) \leq x_i] \quad A_j = [x_j(\xi) \leq x_j, \ x_j(\xi) \leq x_j] = A_i \cap A_j = A_iA_j$$

$x_i$ and $x_j$ are independent if:


Multivariate Probability Distributions

Multivariate cumulative distribution function (CDF), for $x, y$ continuous or discrete:

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

Multivariate probability mass function (PMF), for $x, y$ discrete:

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

Multivariate probability density function (PDF), for $x, y$ continuous:
\[
\frac{\partial^2 F_{xy}(x, y)}{\partial x \partial y}
\]

If \(x\) and \(y\) are independent:

\[
F_{xy}(x, y) = P[x \leq x]P[y \leq y] = F_x(x)F_y(y)
\]

\[
p_{xy}(x_i, y_j) = p_x(x_i)p_y(y_j)
\]

\[
f_{xy}(x, y) = f_x(x)f_y(y)
\]

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 \(x\) and \(y\) is frequently described with the covariance and correlation:

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

\[
\text{Correl}(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}} = \frac{\text{Cov}(x, y)}{\text{Std}(x)\text{Std}(y)}
\]

Uncorrelated \(x\) and \(y\): \(\text{Cov}(x, y) = \text{Correl}(x, y) = 0\)

Independence implies uncorrelated (but not necessarily vice versa)

Examples

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

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

\(a_x = E(x), a_y = E(y), \text{ Correl}(x, y) = 0\)
Two dependent normally distributed random variables (parameters $\mu_x$, $\mu_y$, $\sigma_x$, $\sigma_y$, and $\rho$):

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

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

$\mu$ = vector of means $= [E(x) \ E(y)]'$

$C$ = 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($x$), $\sigma_y$ = Std($y$), $\rho$ = Correl($x,y$)

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

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

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 $\rho$ to examine its effect on the scatter. Consider $\rho = 0., 0.5, 0.9$. Use `subplot` to put plots for all $3 \rho$ values on one page.

---

MIT Copyright 2003 Massachusetts Institute of Technology
Last modified Oct. 8, 2003

3