EECS 126: Probability & Random Processes
Fall 2021

Basic Probability

Shyam Parekh

General Framework

- Probability Space = {Sample Space Ω, Events, Probability Measure P}
  - For an event A, $0 \leq P(A) \leq 1$
  - $P(\phi) = 0, P(\Omega) = 1$
  - If the events A_1, A_2, A_3 ... are disjoint, $P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$

Borel-Cantelli Theorem

- Let $\{A_n, \mathrm{n} \geq 1\}$ be a collection of events such that $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n \text{ infinitely often}) = 0$.

Independence

- Consider events $\{A_j, j \epsilon J\}$ where J is a set of some positive integers.
  - Pairwise independence: For any pair $j, k \ \epsilon \ J, P\left(A_j \cap A_k\right) = P(A_j)P(A_k)$.
  - Mutual independence: $P(\bigcap_{j \epsilon K} A_j) = \prod_{j \epsilon K} P(A_j)\,, \forall \ \ finitie \ K \subset J.$
- Pairwise independence $\nRightarrow$ Mutual independence

Converse of Borel-Cantelli Theorem

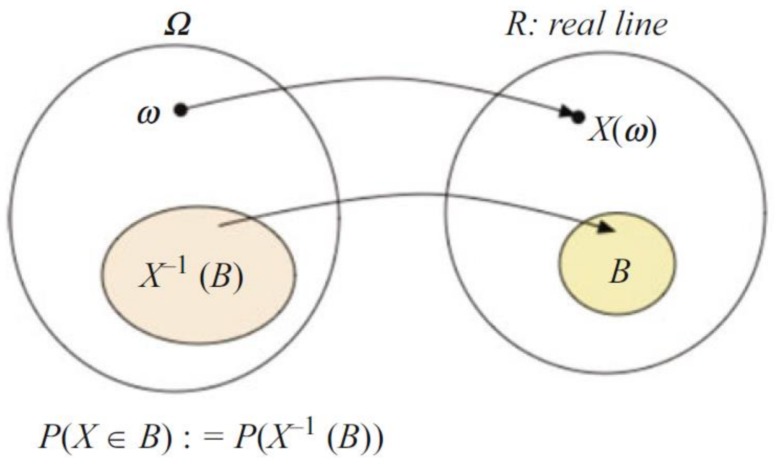
- Let $\{A_n, \mathrm{n} \geq 1\}$ be a collection of mutually independent events such that $\sum_{n=1}^{\infty} P(A_n) = \infty$, then $P(A_n \text{ infinitely often}) = 1$.

Conditional Probability

- Let A and B be two events, and assume $P(B) > 0$.
  - Conditional probability $P[A|B] \coloneqq P(A \cap B)/P(B)$
- If A and B are independent, $P[A|B] = P(A)$
- $P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P[A_2|A_1]P[A_3|A_1 \cap A_2] \ldots P[A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}]$, if $P(A_1 \cap A_2 \cap \cdots A_{n-1}) > 0$.
- Bayes' Rule:
  - Let $A, B_1, \ldots, B_n$ be the events where B_i's are disjoint and $\bigcup_{i=1}^{n} B_i = \Omega$ (sample space).
  - Then, $P[B_i|A] = P[A|B_i]P(B_i) \,/\, \sum_{j=1}^{n} P[A|B_j]P(B_j)$
  - This uses $\sum_{j=1}^{n} P(A \cap B_j) = P(A)$

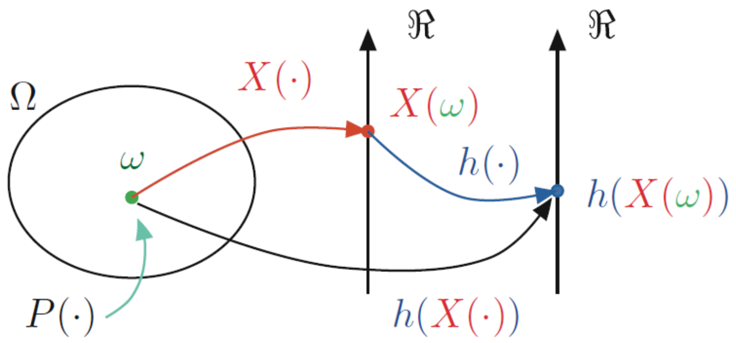
Random Variable

- A Random Variable (RV) X is a function $X: \Omega \to R$.
- For a $B \subset R$, $P(X \epsilon B) = P(X^{-1}(B))$, where $X^{-1}(B) = \{\omega \epsilon \Omega : X(\omega) \epsilon B\}$.
- Cumulative Distribution Function (CDF) of X: $F_X(x) \coloneqq P(X \epsilon (-\infty, x]) = P(X \leq x)$
  - F_X is non-decreasing and right-continuous.
  - $F_X(x) \to 0$ as $x \to -\infty$ and $F_X(x) \to 1$ as $x \to \infty$.



Discrete Random Variable

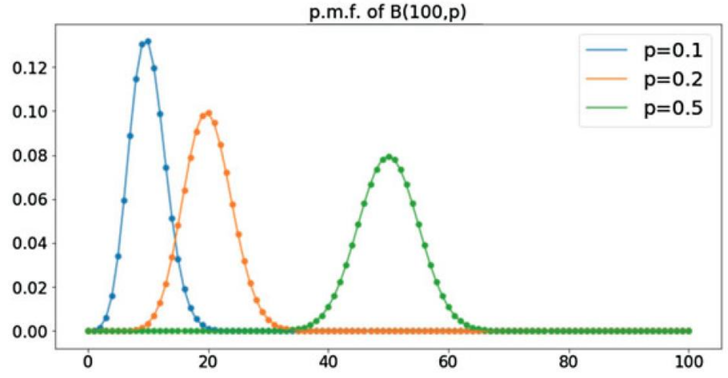
- $X \equiv \{(x_n, p_n), n = 1, \ldots, N\}$ where $p_n = P(X = x_n)$.
  - N can be infinite.
  - $\{(x_n, p_n), n = 1, \ldots, N\}$ is called the Probability Mass Function (PMF).
- $E(X) = \sum_{n=1}^{N} x_n p_n$
  - With $N = \infty$, expectation is undefined if both the sums of positive and negatives terms are infinite.
- Function of Random Variable: $\{(h(x_n), p_n), n = 1, \ldots, N\}$.
  - For defining PMF, we merge identical values and add their probabilities.

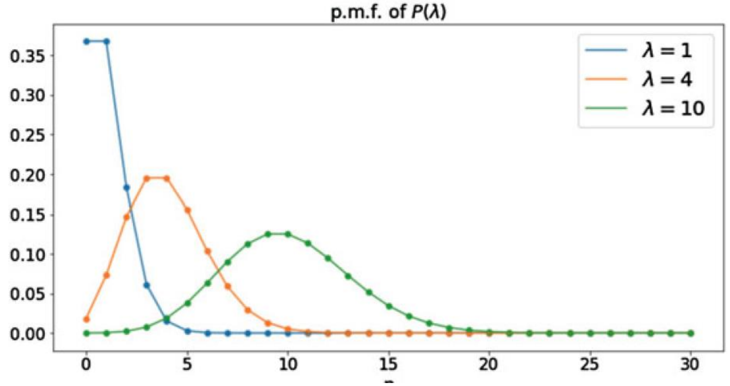


- $E\big(h(X)\big) = \sum_{n=1}^{N} h(x_n) p_n$
- $var(X) = \mathrm{E}\left(\left(X - E(X)\right)^2\right) = \mathrm{E}(X^2) - [E(X)]^2 = {\sigma_X}^2$
- Coefficient of Variation: CV $= \sigma_X / E(X)$
- Refer to the last lecture for the properties of expectation and variance.

Important Discrete Random Variables

- Bernoulli: $X =_D B(p)$ with $p \epsilon [0,1]$
  - $X \equiv \{(0, 1-p), (1, p)\}$
  - $E(X) = p, var(X) = p(1-p)$
  - Canonical example: Models coin flip with 1 for “heads” and 0 for “tails”
- Geometric: $X =_D G(p)$ with $p \epsilon [0,1]$
  - $P(X = n) = (1-p)^{n-1}p, n \geq 1$
  - $E(X) = 1/p, var(X) = (1-p)p^{-2}$
  - X is memoryless: $P[X > m + n \mid X > m] = P(X > n) = (1-p)^n, \mathrm{m}, \mathrm{n} \geq 1$
  - Canonical example: Models # of coin flips until the first “heads”
- Binomial: $X =_D B(N, p)$ with $p \epsilon [0,1]$ and $\mathrm{N} \geq 1$
  - $P(X = n) = \binom{N}{n} p^n (1-p)^{N-n}, n = 0, 1, \dots, N$
  - $E(X) = Np, var(X) = Np(1-p)$
  - Canonical example: Models # of “heads” in N coin flips
- Poisson: $X =_D P(\lambda)$ with $\lambda > 0$
  - $P(X = n) = \frac{\lambda^n}{n!} e^{-\lambda}, n \geq 0$
  - $E(X) = \lambda, var(X) = \lambda$
  - Canonical example: Models # arrivals in a given interval





Multiple Discrete Random Variables

- Consider a pair of RVs (X, Y)
- Joint Probability Mass Function (JPMF): $p_{i,j} = P\left(X = x_i, Y = y_j\right) \forall\, i, j$
  - The number of possible values for X and/or Y can be infinite.
- Marginal PMF from JPMF: $P(X = x_i) = \sum_j P(X = x_i, Y = y_j)$
- RVs X, Y are independent if $P(X = x, Y = y) = P(X = x)P(Y = y) \;\; \forall\, x, y$
- For h: $R^2 \to R, E\left(h(X,Y)\right) \coloneqq \sum_i \sum_j h(x_i, y_j) p_{i,j}$
- As before, $cov(X,Y) = E(\left(X - E(X)\right)\left(Y - E(Y)\right) = E(XY) - E(X)E(Y)$
  - Uncorrelated vs. positively/negatively correlated
- Theorem B.4:
  - Independent RVs are uncorrelated.
  - Converse is not true.
  - The variance of a sum of uncorrelated RVs is the sum of their variances.

Multiple Discrete Random Variables (Cont’d)

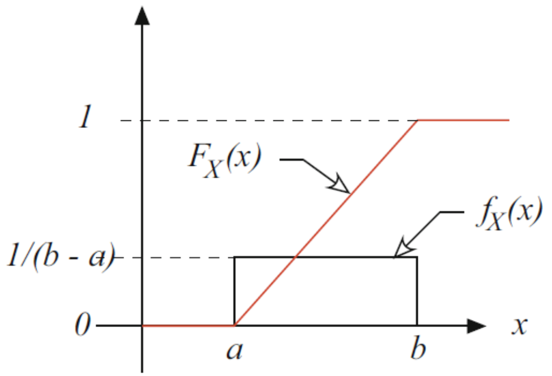
- Conditional PMF: $P\big(Y = y_j \,\big|\, X = x_i\big) = P(X = x_i, Y = y_j)/P(X = x_i)$
- Conditional expectation: $E[Y \mid X = x_i] = \sum_j y_j P\big[Y = y_j \,\big|\, X = x_i\big]$
  - We define $E[Y \mid X]$ as an RV that takes the value $E[Y \mid X = x_i]$ when $X = x_i$.
  - Observe $E[Y \mid X]$ is a function $g(X)$ with $g(x_i) = E[Y \mid X = x_i]$.
- Theorem B.5:
  - $E[E[Y \mid X]] = \mathrm{E[Y]}$
  - $E[h(X)Y \mid X] = h(X)E[Y \mid X]$
  - $E[Y \mid X] = E(Y)$ if X and Y are independent.
- Conditional expectation of a function of an RV
  - $E[h(Y) \mid X = x_i] = \sum_j h(y_j)P\big[Y = y_j \,\big|\, X = x_i\big]$
- Linearity of conditional expectation
  - $E[h_1(Y) + h_2(Y) \mid X] = E[h_1(Y) \mid X] + E[h_2(Y) \mid X]$

General Random Variables

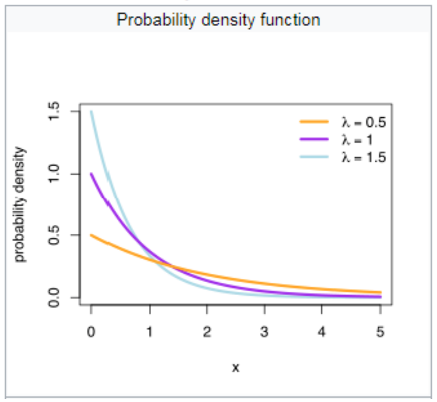
- Continuous range of possible values.
- Cumulative Distribution Function (CDF) of an RV X is defined as $F_X(\mathrm{x}) = \mathrm{P}(\mathrm{X} \leq x)$.
- Probability Density Function (PDF) of an RV X is defined as $f_X(\mathrm{x}) = \frac{d}{dx} F_X(x)$ if the derivative exists.
- Observations:
  - $P(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx.$
  - $f_X(x)dx = P(X \epsilon (x, x + dx]).$

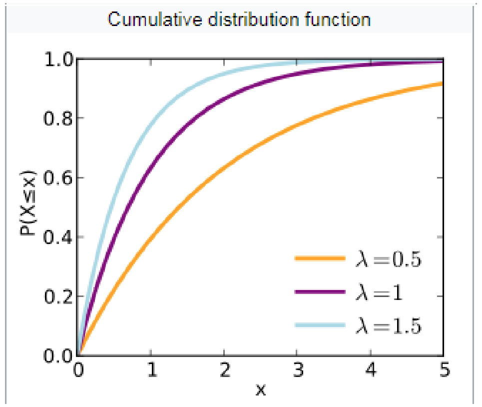
Examples of General Random Variables

- Uniform: $X =_D U[a,b]$ with $a < b$
  - PDF $f_X(\mathrm{x}) = \frac{1}{(b-a)} 1\{a \leq x \leq b\}$
  - CDF $F_X(\mathrm{x}) = \max\{0, \min\left\{1, \frac{x-a}{b-a}\right\}\}$
  - $E(X) = \frac{a+b}{2}, var(X) = \frac{(b-a)^2}{12}$



- Exponential: $X =_D Exp(\lambda)$ with $\lambda > 0$
  - PDF $f_X(\mathrm{x}) = \lambda e^{-\lambda x} 1\{x \geq 0\}$
  - CDF $F_X(x) = 1 - e^{-\lambda x}, \mathrm{x} \geq 0$
  - X is memoryless: $P[X > x + y |\, X > x] = P(X > y) = e^{-\lambda y}, x, y > 0$
  - $E(X) = \lambda^{-1}, \mathrm{var(X)} = \lambda^{-2}$

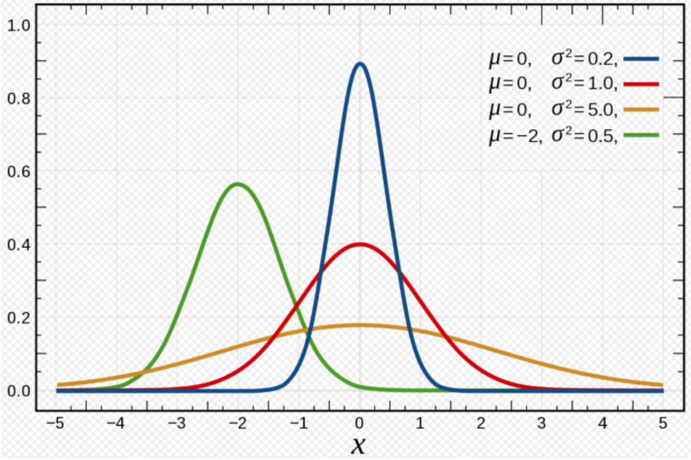




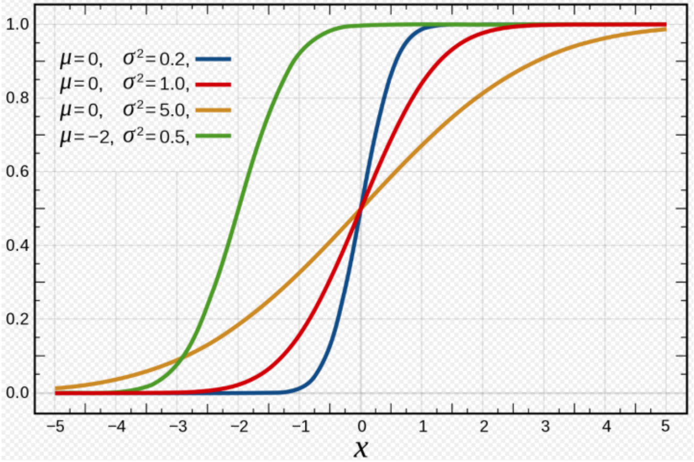
Examples of General Random Variables (Cont'd)

- Gaussian/Normal: $X \;=_D\; N(\mu, \sigma^2)$
  - PDF $f_X(\mathrm{x}) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2)$
  - CDF $F_X(\mathrm{x}) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{1}{2}(\frac{u-\mu}{\sigma})^2)\mathrm{du}$
  - $E(X) = \mu, var(X) = \sigma^2$

Probability density function



Cumulative distribution function



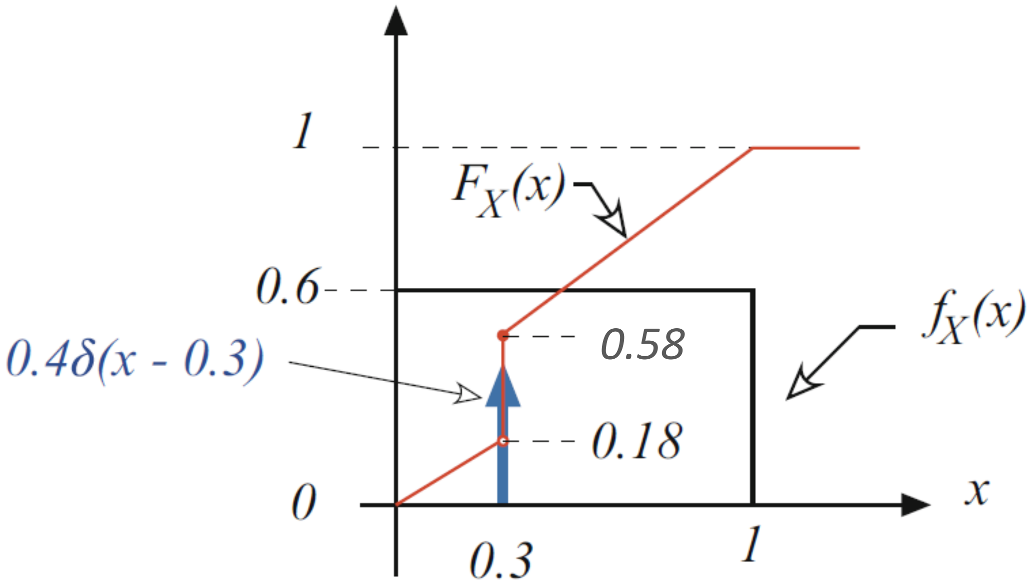
Expectation of General Random Variables

- $E\big(h(X)\big) = \int_{-\infty}^{\infty} h(x) dF_X\,(\mathrm{x}) = \int_{-\infty}^{\infty} h(x) f_X(x) dx$
  - Example: $X\ =_D\ U[0,1], E\big(X^k\big) = \int_0^1 x^k\,\mathrm{dx} = 1/(\mathrm{k}+1)$
- In general, $X_n \to \mathrm{X} \nRightarrow E(X_n) \to E(X)$
  - However, Dominated Convergence Theorem (DCT) and Monotone Convergence Theorem (MCT) provide sufficient conditions for this to happen.
- Theorem: Let $X \geq 0$ be a non-negative RV with $E(X) < \infty$. Then,

  $E(X) = \int_0^{\infty} P(X > x) dx.$
  - Proof relies on DCT.
  - Expectation is integral of Complementary CDF (CCDF).

Mixed Random Variable Example

- With probability 0.4, $X = 0.3$ and is uniformly distributed in $[0, 1]$ with probability 0.6.
- Find $CDF, PDF, E(X)\ and\ var(X)$.

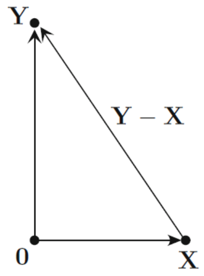


Multiple General Random Variables

- For RVs X and Y, Joint CDF (JCDF) $F_{X,Y}(\mathrm{x},\mathrm{y}) = \mathrm{P}(\mathrm{X} \leq x, Y \leq y)$.
  - If exists, Joint PDF (JPDF) $f_{X,Y}(x,y)$ satisfies $F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(\mathrm{x},\mathrm{y})\mathrm{dxdy}$.
  - Interpret $f_{X,Y}(\mathrm{x},\mathrm{y})\mathrm{dxdy} = \mathrm{P}(\mathrm{X} \in (x, x+dx], Y \in (y, y+dy])$.
  - Example: $f_{X,Y}(x,y) = \frac{1}{\pi}1\{x^2 + y^2 \leq 1\}, \mathrm{x},\mathrm{y} \in R$ places (x,y) uniformly over a unit circle.
- For h: $R^2 \to R, E\big(h(X,Y)\big) \coloneqq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y) f_{X,Y}(x,y)\,\mathrm{dxdy}$.
- If exist, JPDF's of independent RVs X and Y satisfy $f_{X,Y}(x,y) = f_X(x) f_Y(y)\ \forall\ x, y \in R$.
- For independent RVs X and Y, given two functions $g, h\colon R \to R$, RVs $g(X)$ and $h(Y)$ are independent.
- Minimum and maximum of independent RVs X and Y:
  - With $V = \min(X,Y)$, $P(V > v) = P(X > v, Y > v) = P(X > v)P(Y > v)$.
  - With W$= max(X,Y)$, $P(W \leq w) = P(X \leq w, Y \leq w) = P(X \leq w)P(Y \leq w)$.
- Sum of independent RVs X and Y: $Z = X + Y$.
  - $f_Z(z)dz = \int_{-\infty}^{\infty} f_X(\mathrm{x}) f_Y(z-x)dxdz \Rightarrow f_Z(z) = \int_{-\infty}^{\infty} f_X(\mathrm{x}) f_Y(z-x)dx$
  - $f_Z(\mathrm{z}) = f_X * f_Y(\mathrm{z})$ (i.e., f_Z is convolution of f_X and f_Y).
- Conditional PDF and expectation.
  - $f_{Y|X}[y \mid x] = f_{X,Y}(x,y)/f_X(x)$ if $f_X(x) > 0$.
  - $E[h(Y) \mid X = x] = \int_{-\infty}^{\infty} h(y)\, f_{Y|X}[y \mid x]dy$.

Random Vectors

- Let $\boldsymbol{X} = (X_1, \dots, X_n)' = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ be a random vector in R^n where the RVs are defined over a common probability space.
- Given two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, we define
  - $E(\boldsymbol{X}) = (\mathrm{E}(X_1), \dots, \mathrm{E}(X_n))'$
  - $\Sigma_{\boldsymbol{X}} = \mathrm{E}\left(\left(\mathbf{X} - \mathrm{E}(\mathbf{X})\right)\left(\mathbf{X} - \mathrm{E}(\mathbf{X})\right)'\right)$ (an nxn matrix of $E\left(\left(X_i - E(X_i)\right)\left(X_j - E(X_j)\right)\right)$)
  - $cov(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{E}\left(\left(\mathbf{X} - \mathrm{E}(\mathbf{X})\right)\left(\mathbf{Y} - \mathrm{E}(\boldsymbol{Y})\right)'\right)$ (an nxn matrix of $E\left(\left(X_i - E(X_i)\right)\left(Y_j - E(Y_j)\right)\right)$)
  - Note $\Sigma_{\boldsymbol{X}} = cov(\boldsymbol{X}, \boldsymbol{X}) =: cov(\boldsymbol{X})$
  - It can be shown that $cov(A\boldsymbol{X} + \boldsymbol{a}, B\boldsymbol{Y} + \boldsymbol{b}) = Acov(\boldsymbol{X}, \boldsymbol{Y})B'$
- Given two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, we define their inner product $< \boldsymbol{X}, \boldsymbol{Y} > := E(\boldsymbol{X}'\boldsymbol{Y}) = \sum_{i=1}^{n} E(X_i Y_i)$ (Note: Textbook defines $\boldsymbol{X} \perp \boldsymbol{Y}$ if $E(\boldsymbol{X}\boldsymbol{Y}') = \boldsymbol{0}$, where both sides are nxn matrices.)
  - Define $||\boldsymbol{X}||^2 =< \mathbf{X}, \mathbf{X} > = \sum_{i=1}^{n} E(X_i^2)$
  - If $< \boldsymbol{X}, \boldsymbol{Y} > = 0$, we say X, Y are orthogonal or $\boldsymbol{X} \perp \boldsymbol{Y}$.
  - If $< \boldsymbol{X}, \boldsymbol{Y} > = 0, ||\boldsymbol{Y} - \boldsymbol{X}||^2 = ||\boldsymbol{X}||^2 + ||\boldsymbol{Y}||^2$ (“Pythagoras’ Theorem”)



- If random vectors $\boldsymbol{X}^i \perp \boldsymbol{Y}^j$ (where $\boldsymbol{U}^i$ has i^{th} component as the original U_i and all other components set to 0) for $\forall\, i, j$, $E(\boldsymbol{X}\boldsymbol{Y}') = \boldsymbol{0}$. (Both sides are nxn matrices.)
  - If $E(\boldsymbol{X}) = \boldsymbol{0}$, $cov(\boldsymbol{X}, \boldsymbol{Y}) = \boldsymbol{0}$.
  - This implies $\boldsymbol{X} \perp \boldsymbol{Y}$, and $||\boldsymbol{Y} - \boldsymbol{X}||^2) = (||\boldsymbol{X}||^2) + (||\boldsymbol{Y}||^2$.

Density of a Function of RVs

- Suppose a RV X has PDF $f_X(\mathrm{x})$ and $Y = aX + b$. Then,

  $f_Y(y) = \frac{1}{|a|} f_X(\mathrm{x})$ where $ax + b = y$.

- Suppose a random vector $\boldsymbol{X}$ has JPDF $f_{\boldsymbol{X}}(\boldsymbol{x})$, and $\boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b}$ where A is a nonsingular matrix. Then,

  $f_{\boldsymbol{Y}}(\boldsymbol{y}) = \frac{1}{|A|} f_{\boldsymbol{X}}(\mathbf{x})$ where $A\boldsymbol{x} + \boldsymbol{b} = \boldsymbol{y}$, and $|A|$ is the absolute value of the determinant of A.

- Suppose $\boldsymbol{Y} = g(\boldsymbol{X})$ where $\boldsymbol{X}$ has density $f_{\boldsymbol{X}}(\boldsymbol{x})$. Then,

  $f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_i \frac{1}{|J(\boldsymbol{x_i})|} f_{\boldsymbol{X}}(\boldsymbol{x_i})$ where the sum is over all $\boldsymbol{x_i}$ such that $g(\boldsymbol{x_i}) = \boldsymbol{y}$ and $|J(\boldsymbol{x_i})|$ is the absolute value of Jacobian evaluated at $\boldsymbol{x_i}$.

  – Recall $J_{i,j}(\mathbf{x}) = \frac{\partial}{\partial x_j} g_i(\boldsymbol{x})$.