Probability and Random Variables

Gaurav Sharma

Lecture Objectives

- Review probability and random variables
- Pre-requisite for discussion in rest of course

Probability

- Two philosophical viewpoints:
  - Frequentist
    - Probability of A: Fraction of outcomes A
    - Parameters are fixed and data are a repeatable random sample
    - n experiments, n_A have outcome A
    - $\lim_{n\to\infty} n_A/n = p_A$
  - Bayesian
    - Probability of A = Degree of belief that A is true
      - Suitably calibrated to betting odds
    - Unknown quantities treated probabilistically
      - data are fixed and known, parameters unknown also modeled probabilistically
    - Example: Probability that life came to earth on an asteroid
- We will lean toward the Bayesian camp

Probability Space

- Three elements:
  - Set of outcomes Ω
  - Events for which probability is assigned $\mathscr{B}$
  - A measure of probability P()
- Discrete Ω
  - Any subset of outcomes defines an event
- Non discrete Ω
  - Technical constraints "measurable"
  - Most events we think of non-problematic

Probability Space Example

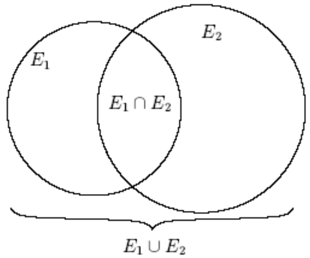
Concept	Example
• Random Experiment	• Throw of a Dice
• Outcomes	• 1, 2, 3, 4, 5, 6
• Sample Space	• {1, 2, 3, 4, 5, 6}
• Events	• {1}, {2,3}, {1,3,5},….
• Probability	• P({2,3})= 2/6 = 1/3
	• P({1}) = 1/6, ….

Probability Axioms

- E – Event
  - $0 \leq P(E) \leq 1$
  - $P(E^c) = 1 - P(E)$
  - $P(\Omega) = 1$
  - Disjoint Events E_i: $\mathscr{P}\left(\bigcup_i E_i\right) = \sum_i \mathscr{P}(E_i).$
- Additional restrictions on what is allowable as a event in non-discrete case

Probability Properties I

$$\mathscr{P}(E^c) = 1 - \mathscr{P}(E^c)$$
$$\mathscr{P}(\phi) = 0$$
$$\mathscr{P}(E_1 \cup E_2) = \mathscr{P}(E_1) + \mathscr{P}(E_2) - \mathscr{P}(E_1 \cap E_2)$$



$$\mathscr{P}(\text{Odd number}) = 1 - \mathscr{P}(\text{Even number})$$

$$E_1 = \text{Multiple of } 2,\ \mathscr{P}(E_1) = \frac{1}{2}$$
$$E_2 = \text{Multiple of } 3,\ \mathscr{P}(E_2) = \frac{1}{3}$$
$$\mathscr{P}(E_1 \cup E_2) = P(\{1, 2, 3, 4, 6\}) = \frac{2}{3}$$
$$\mathscr{P}(E_1 \cap E_2) = P(\{6\}) = \frac{1}{6}$$
$$\frac{2}{3} = \frac{1}{2} + \frac{1}{3} - \frac{1}{6}$$

Conditional Probability

<u>Conditional Probability</u>

$$\mathscr{P}(E_1|E_2) = \text{Probability of } E_1 \text{ given } E_2 = \frac{\mathscr{P}(E_1 \cap E_2)}{\mathscr{P}(E_2)}$$

E_2 forms new universe

Bayes Rule

$$E_1 = \{5\}\ ,\ E_2 = \text{throw} \geq 3 = \{\ 3,\ 4,\ 5,\ 6\ \}$$

$$\mathscr{P}(\{5\}|\{3,4,5,6\}) = \frac{1}{4} = \frac{\frac{1}{6}}{\frac{4}{6}} = \frac{\mathscr{P}(\{5\})}{\mathscr{P}(\{3,4,5,6\})}$$

$\{3,4,5,6\}$ is new "universe"

Note $\mathscr{P}(\{3,4,5,6\}|\{5\}) = 1$

G. Sharma

Statistical Independence

$$\underline{\text{Statistical Independence}}$$

$$\mathscr{P}(E_1 \cap E_2) = \mathscr{P}(E_1)\mathscr{P}(E_2)$$

$$\Leftrightarrow \mathscr{P}(E_1|E_2) = \mathscr{P}(E_1)$$

$$E_2 \text{ overlaps same fraction of } E_1 \text{ as of } \Omega$$

$$E_1 = \text{throw} \geq 3 = \{3, 4, 5, 6\}$$

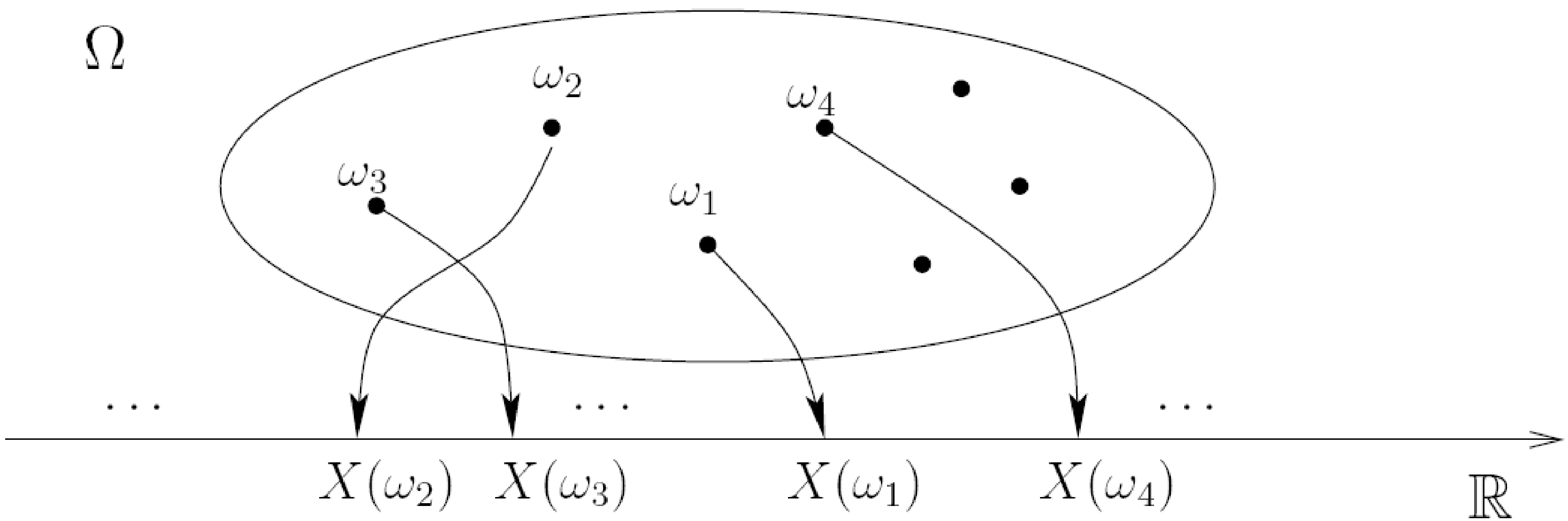
$$E_2 = \text{Odd Throw} = \{1, 3, 5\},\ E_1 \cap E_2 = \{3, 5\}$$

$$\mathscr{P}(\{3, 4, 5, 6\} = \tfrac{4}{6} = \tfrac{2}{3},\ \mathscr{P}(\{1, 3, 5\}) = \tfrac{3}{6} = \tfrac{1}{2}$$

$$\mathscr{P}(\{3, 5\}) = \tfrac{2}{6} = (\tfrac{2}{3})(\tfrac{1}{2}) = \mathscr{P}(\{3, 4, 5, 6\})\mathscr{P}(\{1, 3, 5\})$$

Random Variable

- A real number for each outcome of a random experiment

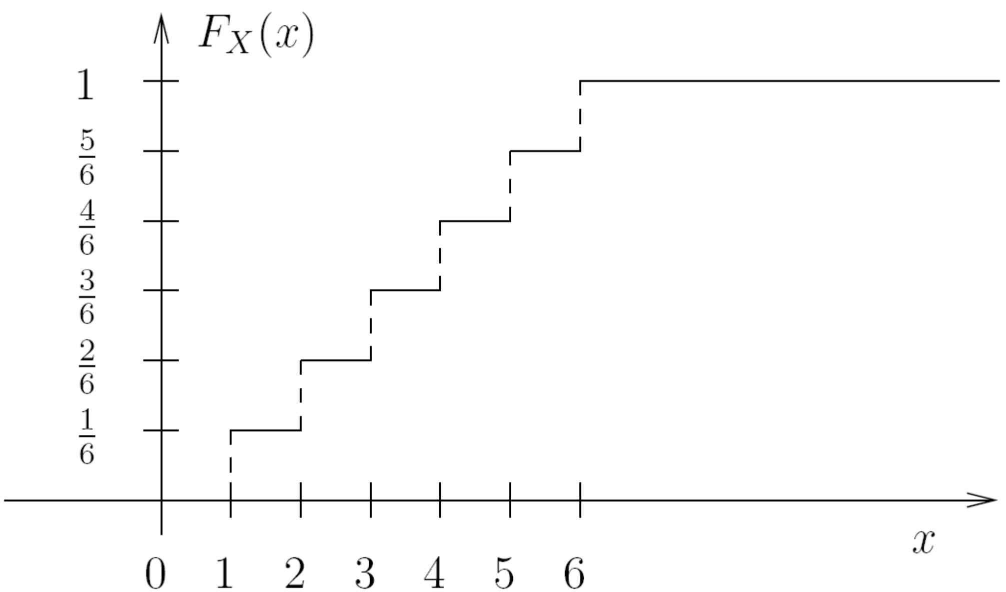


Cumulative Distribution Function

- Cumulative Distribution Function

$$F_X(x) = \mathscr{P}(X \leq x)$$

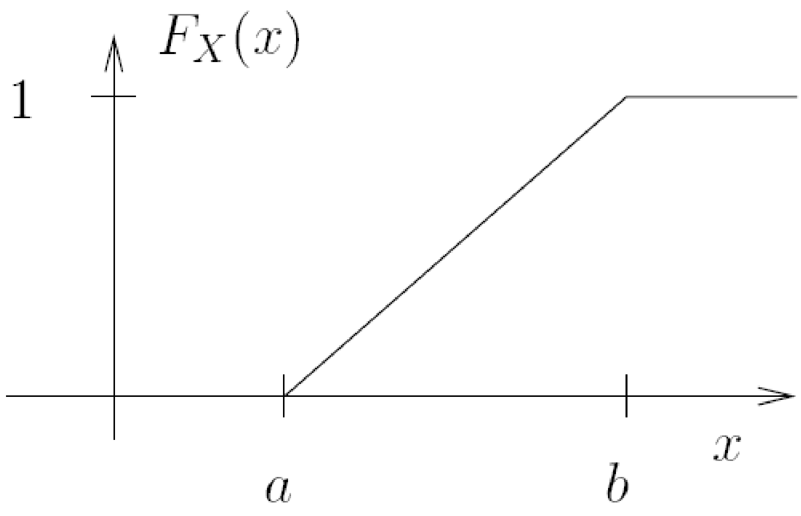
- Dice throw



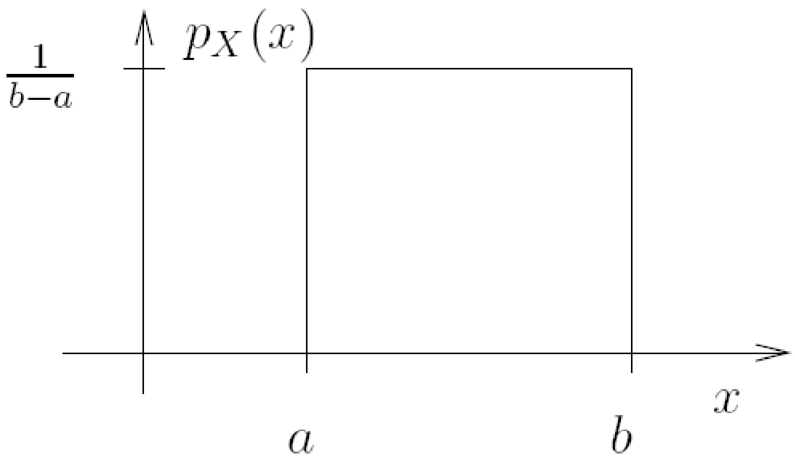
Probability Density Function

- PDF $$p_X(X) = \frac{d}{dx} F_X(x)$$

- Example: Uniform Random Variable



CDF



PDF

Statistical Averages (R.V.)

- Mean/Expected Value

$$m_X = E[X] \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x p_X(x) dx$$

- n^{th} Moment:

$$m_X^{(n)} = E[X^n] \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} x^n p_X(x) dx$$

- Variance/Standard Deviation:

$$\sigma_X^2 = E[(X - m_X)^2] \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} (x - m_X)^2 p_X(x) dx = E[X^2] - m_X^2$$

Characteristic Function

$$\psi_X(\nu) \stackrel{\text{def}}{=} E[e^{j\nu X}] = \int_{-\infty}^{\infty} p_X(x) e^{j\nu x} dx$$

- Fourier transform of pdf
- Utility: Computation of moments

$$m_X^n = \frac{1}{j^n} \frac{d^n \psi_X(\nu)}{d\nu^n} \Big|_{\nu=0}$$

Gaussian Random Variable

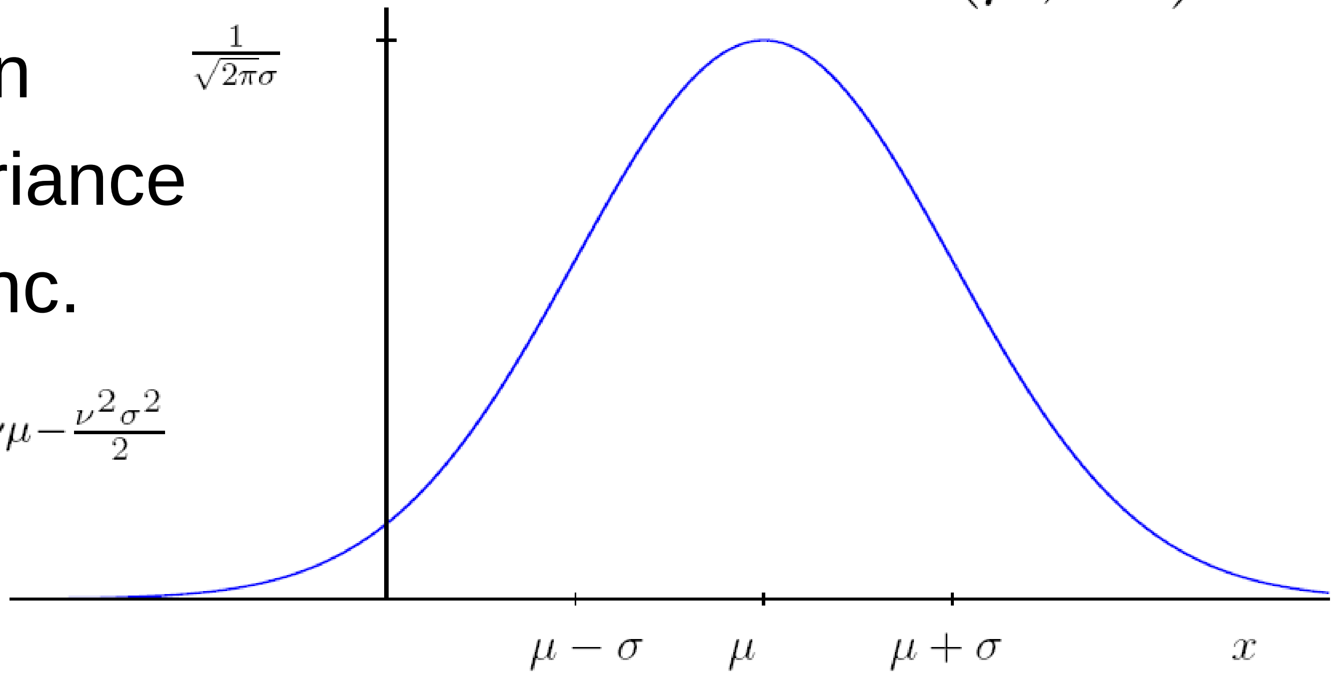
- X: with pdf

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mathcal{N}(\mu, \sigma^2)$$

  - μ - Mean
  - σ^2 – Variance
  - Char func.

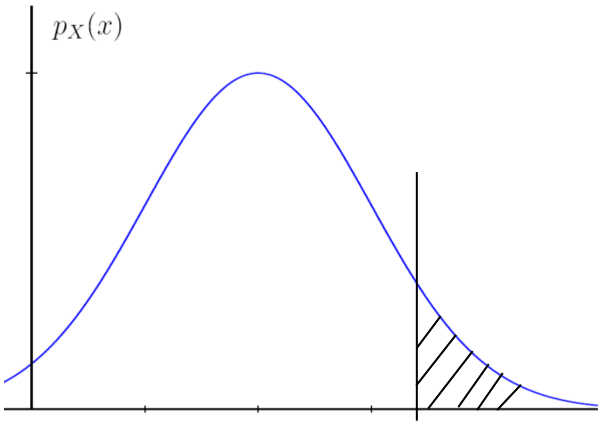
$$\psi_X(\nu) = e^{j\nu\mu - \frac{\nu^2\sigma^2}{2}}$$



Normalized Gaussian RV

$$Y \sim \mathcal{N}(\mu, \sigma^2)$$

$$X = \frac{Y - \mu}{\sigma}$$ is $$\mathcal{N}(0,1)$$

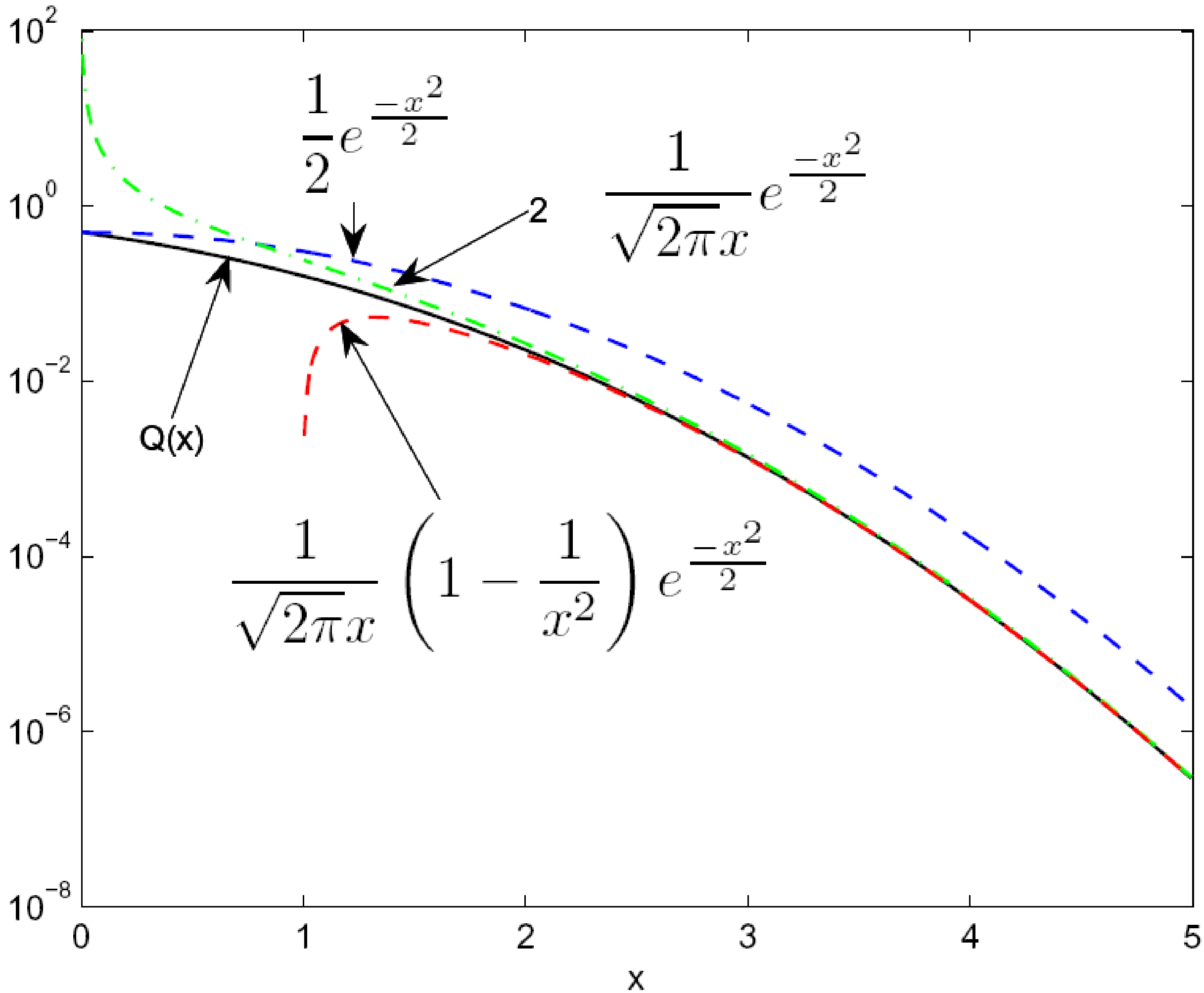


– Normalized Gaussian RV

– Tail probability: Q-function

$$Q(x) = \mathscr{P}(X > x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Q-Function Bounds



Function of a RV

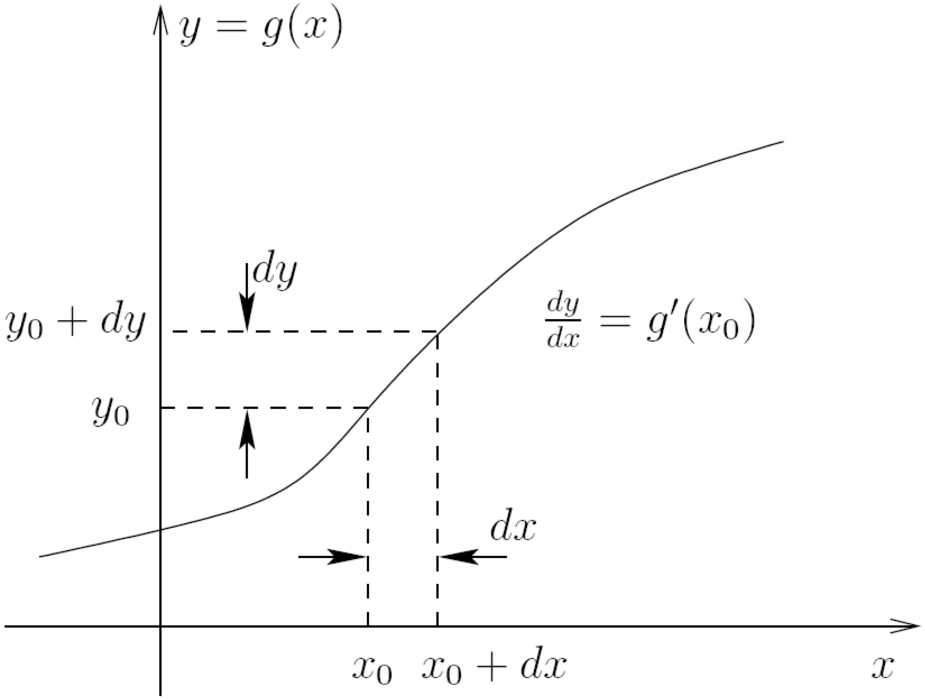
$$Y = g(X)$$

- A new random variable
  - fully characterized by X
- CDF:

$$F_Y(y) = \int_{\{x:g(x)\leq y\}} p_X(x)dx$$

Functions of a RV II

- Non-decreasing function



$$p_Y(y_0) = \frac{p_X(x_0)}{\frac{dy}{dx}} = \frac{p_X(x_0)}{g'(x_0)}$$

$$p_Y(y_0)dy = \mathscr{P}(\{y_0 \le Y < y_0 + dy\}) = \mathscr{P}(\{x_0 \le X < x + dx\}) = p_X(x_0)dx$$

Functions of a RV III

- Countable number of solutions to $g(x)=y$, Say x_i
  - $g'(x_i)$ non-zero at each x_i

$$p_Y(y) = \sum_i \frac{p_X(x_i)}{|g'(x_i)|}$$

Example: RV Transformation I

$$X \sim \mathcal{N}(0,1)$$

$$Y = aX + b$$

$$p_Y(y) = \frac{p_X\left(\frac{y-b}{a}\right)}{|a|} = \frac{1}{\sqrt{2\pi}|a|} e^{-\frac{(y-b)^2}{2|a|^2}}$$

Example: RV Transformation II

- X uniform over $[-\pi,\pi]$

$$f_X(x) = \begin{cases} \frac{1}{2\pi} & -\pi \leq x < \pi \\ 0 & \text{otherwise} \end{cases}$$

- $Y = X^2$
  - $y=x^2$ has two solutions +/- $\sqrt{y}$

$$f_Y(y) = \sum_{x_i : g(x_i)=y} \frac{f_X(x_i)}{|g'(x_i)|} = \begin{cases} \frac{1}{2\pi\sqrt{y}} & 0 \leq y < \pi^2 \\ 0 & \text{otherwise} \end{cases}$$

RV Transformation: Application

- Computer generation of Random Variables with desired distributions

- Assignment 1: Addresses this problem

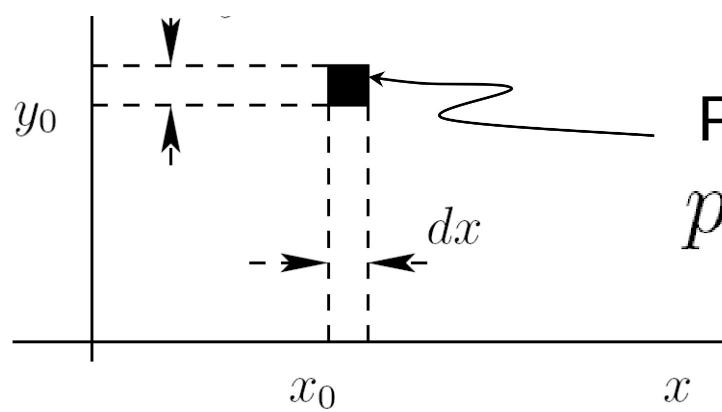
Two Random Variables

- Individual vs Joint Characterization
- Joint cdf

$$F_{XY}(x,y) \;=\; \mathscr{P}(X \leq x, Y \leq y)$$

- Joint pdf

$$p_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$$



Probability

$$p_{XY}(x,y)dxdy$$

Joint/Marginal PDFs

- Marginals integral of joint pdfs

$$p_X(x) = \int_{-\infty}^{\infty} p_{XY}(x, y)dy.$$

$$p_Y(y) = \int_{-\infty}^{\infty} p_{XY}(x, y)dx.$$

Conditional PDF, Independence

$$p_{Y|X}(y|x) = \begin{cases} \frac{p_{XY}(x,y)}{p_X(x)} & p_X(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

- Statistically Independent
  - $p_{Y|X}(y|x) = p_Y(y)$
  - Equivalent:

$$p_{XY}(x,y) = p_Y(y)p_X(x)$$

Joint Statistics of 2 RVs

- Correlation: E[XY]
- Covariance

$$\mathrm{cov}(X,Y) = E[(X - m_X)(Y - m_Y)]$$

  – Correlation of “mean-removed” versions

- Correlation coefficient

$$\rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \qquad |\rho_{XY}| \leq 1$$

Uncorrelated and Independent RVs

- Uncorrelated: Covariance =0
  - Equivalent conditions

$$E[XY] = E[X]E[Y]$$
$$\rho_{XY} = 0$$

- Independent implies uncorrelated not vice versa (important exception later)

Transformation of 2 RVs

- $$\begin{aligned} Z &= g(X,Y) \\ W &= h(X,Y) \end{aligned}$$ Transform of X, Y

- PDF of Z,W:

$$p_{ZW}(z,w) = \sum_i \frac{p_{XY}(x_i, y_i)}{|\det\left(J(x_i, y_i)\right)|}$$

  – (x_i,y_i) solutions to w=g(x,y); z= h(x,y)

  – Jacobian: $J(x,y) = \begin{bmatrix} \frac{\partial g(x,y)}{\partial x} & \frac{\partial g(x,y)}{\partial y} \\ \frac{\partial h(x,y)}{\partial x} & \frac{\partial h(x,y)}{\partial y} \end{bmatrix}$

Example

- X,Y independent RVs uniform [0,1]

$$Z = g(X,Y) = X + Y$$
$$W = h(X,Y) = X - Y$$

$$J(x,y) = \begin{bmatrix} \frac{\partial g(x,y)}{\partial x} & \frac{\partial g(x,y)}{\partial y} \\ \frac{\partial h(x,y)}{\partial x} & \frac{\partial h(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$\det\left(J(x,y)\right) = -2$$

Example (contd.)

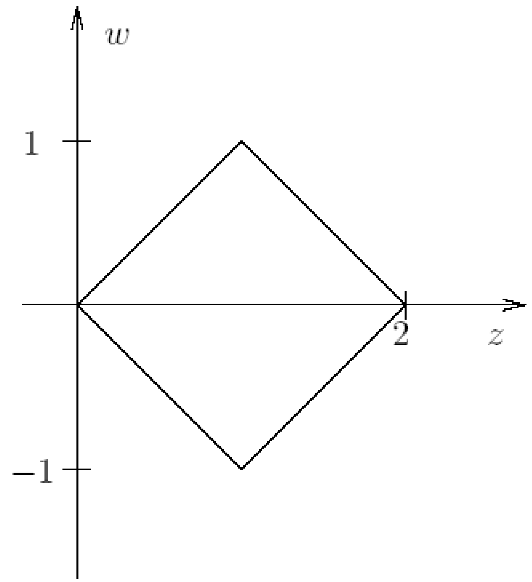
- Unique soln to $x + y = z,\ x - y = w$

$$x_1 = \frac{z+w}{2}, y_1 = \frac{z-w}{2}$$

$$\begin{aligned} p_{ZW}(z,w) &= \frac{p_{XY}(x_1,y_1)}{|\det\left(J(x_1,y_1)\right)|} \\ &= \begin{cases} \frac{1}{2} & 0 \leq z+w \leq 2, 0 \leq z-w \leq 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Example (contd.)

$$p_{ZW}(z, w)$$



- Uncorrelated but dependent
  - Verify !

Example II

- X, Y independent identically distributed (iid) Gaussians $\mathcal{N}(0, \sigma^2)$
- Rectangular to polar conversion

$$
\begin{aligned}
R &= \sqrt{X^2 + Y^2} \stackrel{\text{def}}{=} g(X, Y) \\
\Theta &= \tan^{-1}\left(\frac{Y}{X}\right) \stackrel{\text{def}}{=} h(X, Y)
\end{aligned}
$$

Example II (contd.)

$$\det\left(J(x,y)\right) = \frac{1}{\sqrt{x^2+y^2}}$$

- Solution to $r = \sqrt{x^2+y^2}, \theta = \tan^{-1}\left(\frac{y}{x}\right)$

$$x = r\cos(\theta) \quad y = r\sin(\theta)$$

- Joint pdf of R,θ

$$p_{R\Theta}(r,\theta) = \frac{r}{2\pi\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$$

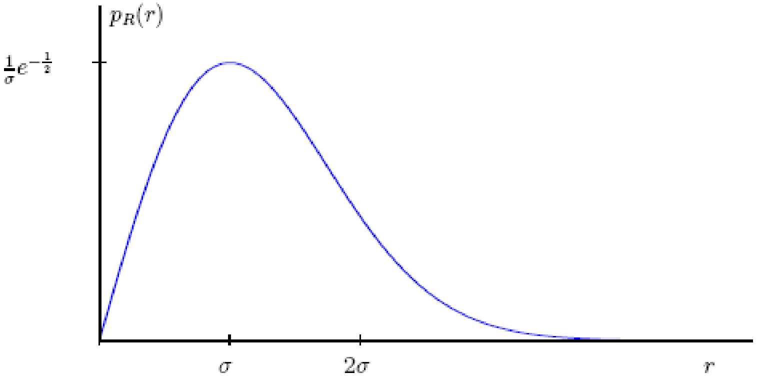
- Independent of θ

Example II (contd.)

- Marginal distribution of R

$$p_R(r) \quad = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \quad r \geq 0.$$

- Rayleigh Distribution
  - Model for amplitude variations in mobile wireless applications



Multiple RVs

- n Random variables: X_1, X_2, … X_n
- Random Vector:

$$\mathbf{X} = [X_1, X_2, \ldots X_n]^T$$

- Characterization
  - Joint cdf: $F_{\mathbf{X}}(\mathbf{x}) = \mathscr{P}\left(\mathbf{X} \leq \mathbf{x}\right)$
  - Joint pdf: $p_{\mathbf{X}}(\mathbf{x}) = \dfrac{\partial^n F_{\mathbf{X}}(\mathbf{x})}{\partial x_1 \partial x_2 \cdots \partial x_n}$

Multiple RV Coordinate Transformation

$$\mathbf{X} = [X_1, X_2, \ldots X_n]^T, \; \mathbf{Y} = [Y_1, Y_2, \ldots Y_n]^T$$

$$\mathbf{Y} = \mathbf{g}(\mathbf{X})$$

$$p_{\mathbf{Y}}(\mathbf{y}) = \sum_i \frac{p_{\mathbf{X}}(\mathbf{x}_i)}{|\det(\mathbf{J}(\mathbf{x}_i))|}$$

– $\mathbf{x}_i$ solution to $\mathbf{g}(\mathbf{x}) = \mathbf{y}$ Jacobian:

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial g_1(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial g_n(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial g_n(\mathbf{x})}{\partial x_n} \end{bmatrix}$$

Jointly Gaussian RVs

- Joint pdf of form

$$p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det \mathbf{C}}} \exp\left(-\frac{(x-\mathbf{m})^T \mathbf{C}^{-1}(x-\mathbf{m})}{2}\right)$$

  – Mean $\mathbf{m} = E[X_1, X_2, \ldots X_n]$

  – Covariance: $\mathbf{C} = E\left[(\mathbf{x}-\mathbf{m})(\mathbf{x}-\mathbf{m})^T\right]$

  – Completely characterized by mean and covariance

Properties of Multivariate Gaussian Random Variables

- Any subset is also jointly Gaussian
- Linear transforms preserve Gaussianity
- Conditional distribution of some given the rest is Gaussian
- Jointly Gaussian + uncorrelated implies independent

G. Sharma

Asymptotic Behavior of Infinite Collections of Random Variables

- Two main laws:
  - Law of large numbers
  - Central Limit Theorem

Law of Large Numbers (LLN)

- X_1, X_2, …. Uncorrelated random variables, common mean m_x, finite variance

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad n = 1, 2, \ldots$$

- Y_n converges to the mean as $n \to \infty$

$$\lim_{n\to\infty} \mathscr{P}\left(|Y_n - m_X| > \epsilon\right) = 0.$$

Central Limit Theorem

- X_1, X_2, …. Independent identically distributed random variables, mean m_x, variance σ_x^2 (finite)

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Converges to $\mathcal{N}(m, \frac{\sigma^2}{n})$ random variable as n $\rightarrow\infty$

Central Limit Theorem Significance

- Noise: often composed of smaller similar noise processes
  – CLT leads to Gaussian distribution
- Gaussian assumption common for several noise sources

Lecture Key Concepts

- Probability space
  - Universe of outcomes, Events, Probabilities
  - Independence
- Random Variables
  - Characterization by cdf/pdf
  - Joint pdf for multiple rvs
  - Transformations
- Law of Large Numbers
- Central Limit Theorem