Chapter 2 Random Variables and Univariate Probability Distributions

2.1 Random Variables and Distribution Functions

Review:

- **Definition: Sigma Algebra (or σ-algebra, Borel field, sigma field).** A sigma algebra, denoted by $\mathcal{B}$, is a **collection of subsets** of S with

(1) $\Phi \in \mathcal{B}$ (the empty set is contained in $\mathcal{B}$).

(2) If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$.

(3) If $A_1, A_2, \cdots, \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$.

- **Definition: Probability Function.** Suppose a random experiment has a sample space S and an associated σ-algebra $\mathcal{B}$. The probability function $P : \mathcal{B} \to [0, 1]$ is a mapping that satisfies the following properties:

(1) $0 \leq P(A) \leq 1$ for any event A in $\mathcal{B}$.

(2) $P(S) = 1$.

(3) If $A_1, A_2, \cdots \in \mathcal{B}$ are mutually exclusive, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

2.1 Random Variables and Distribution Functions

- **Definition: Probability Space.** A *probability space* is a triple $(S, \mathcal{B}, P)$, where
  - S is the sample space corresponding to outcomes of the underlying random experiment.
  - $\mathcal{B}$ is an associated σ-algebra of S. The elements in $\mathcal{B}$ are called events.
  - P is a probability function (probability measure) defined on $(S, \mathcal{B})$.

- It is inconvenient to work with different sample spaces. We need to unify different sample space. Elements of S may be represented by numbers.

2.1 Random Variables and Distribution Functions : R.V.

- **Definition: Random Variable.** A random variable (r.v.), $X(\cdot)$, is a $\mathcal{B}$- measurable mapping (or point function) from the sample space S to the real line $\mathcal{R}$ such that to each outcome $s \in S$ there exists a corresponding unique real number, $X(s)$. The collection of all possible values that the random variable X can take, also called the range of $X(\cdot)$, constitutes a new sample space, denoted as Ω.

- The function $X : S \to \Omega$ need not be a one to one mapping. Thus, it is possible that two basic outcomes $s_1, s_2 \in S$ will deliver the same value for random variable X, $X(s_1) = X(s_2)$.

2.1 Random Variables and Distribution Functions : R.V.

- **Example.** Suppose we throw three fair coins. Then the sample space

$$S = \{TTT,\ TTH,\ THT,\ HTT,\ HHT,\ HTH,\ THH,\ HHH\}.$$

Let $X(\cdot)$ be the number of heads shown up. Then $X(T,T,T) = 0$, $X(T,T,H) = 1$, $X(T,H,T) = 1$, $X(H,T,T) = 1$, $X(H,H,T) = 2$, $X(H,T,H) = 2$, $X(T,H.H) = 2$, $X(H,H,H) = 3$. We have $\Omega = \{0,1,2,3\}$.

- Here, $P(X = 3) = P(A)$, where $A = \{s \in S : X(s) = 3\} = \{HHH\}$ denotes the probability that exactly three heads occur in the experiment.

- X — r.v.; x — realization.

2.1 Random Variables and Distribution Functions : R.V.

- **Remarks.** Suppose we have a sample space with a finite number of basic outcomes $S = \{s_1, \ldots, s_n\}$ and a probability function $P : \ \mathcal{B} \to [0, 1]$, where $\mathcal{B}$ is a σ-field associated with S. Also, we define a random variable $X : S \to \mathcal{R}$ with the range $\Omega = \{x_1, \ldots, x_m\}$, where m may not be the same as n. Then we can define the probability function $P_X : \Omega \to \mathcal{R}$ for the random variable X in the following way:

$$P_X(x_i) \equiv P(X = x_i) = P(\{s \in S : \ X(s) = x_i\}).$$

Here, $P_X(\cdot)$ is an induced probability function on Ω. It is defined in terms of the original probability function $P(\cdot)$. Thus an r.v. X is a function that carries the probability from the original sample space S to a new space Ω of real numbers.

2.1 Random Variables and Distribution Functions : R.V.

- **Question:** What function form $X(\cdot)$ will ensure $C_A \in \mathcal{B}$?

- **Definition: Measurable Function.** A function $X : S \to \mathcal{R}$ is $\mathcal{B}$-measurable (or measurable with respect to the σ-field $\mathcal{B}$ generated from S) if for every real number a, the set $\{s \in S : X(x) \leq a\} \in \mathcal{B}$.

- **Theorem.** Let $\mathcal{B}$ be a σ- algebra associated with sample space S. Let $f(\cdot)$ and $g(\cdot)$ be $\mathcal{B}$-measurable real valued functions, and c be a real number. Then the functions $c \cdot f(\cdot)$, $f(\cdot) + g(\cdot)$, $f(\cdot) \cdot g(\cdot)$ and $|f(\cdot)|$ are also $\mathcal{B}$-measurable.

- Standard functions are measurable and any standard sequence of countable operations on such functions will not destroy measurability.

2.1 Random Variables and Distribution Functions : R.V.

- **Theorem:** Let X and Y be random variables, then $X+Y$ is also a random variable.

- **Proof.** We only need to prove that $X+Y$ is a $\mathcal{B}$-measurable function.

  – For any $a \in \mathbb{R}$,

  $$\{s : (X+Y)(s) < a\} = \cup_{q\in\mathbb{Q}}[\{s : X(s) < a-q\} \cap \{s : Y(s) < q\}] \in \mathcal{B},$$

  where $q \in \mathbb{Q}$ is a rational number.

  – For any $b \in \mathbb{R}$,

  $$\{s : (X+Y)(s) \leq b\} = \cap_{i=1}^{\infty}\{s : (X+Y)(s) < b + \frac{1}{i}\} \in \mathcal{B}.$$

- **Remark:** $\{s \in S : X(s) < x\} \in \mathcal{B}$ for any $x \in \mathbb{R}$ also implies $\mathcal{B}$-measurable.

- **Example: Discrete R.V.** Let's continue the previous example of tossing coins. Suppose we are interested in calculating the probability that $P(0 \leq X \leq 1)$. Denote

$$C = \{s \in S,\ 0 \leq X(s) \leq 1\} = \{TTT, TTH, THT, HTT\}.$$

It follows that

$$P(0 \leq X \leq 1) = P(C) = \frac{1}{2}.$$

- **Question:** How to characterize a random variable X?

2.1 Random Variables and Distribution Functions : CDF

- **Definition: Cumulative Distribution Function (CDF).** *Cumulative Distribution Function* of a random variable X is defined as:

$$F_X(x) = P(X \leq x) \text{ for all } x \in \mathbb{R}.$$

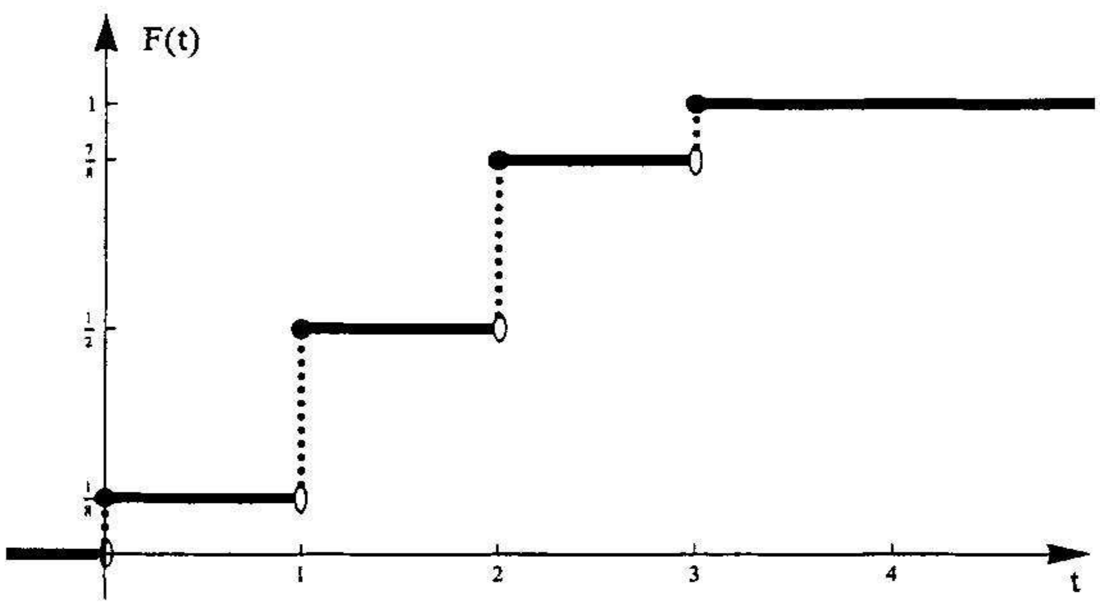

Figure 1: CDF of the example of tossing coin three times.

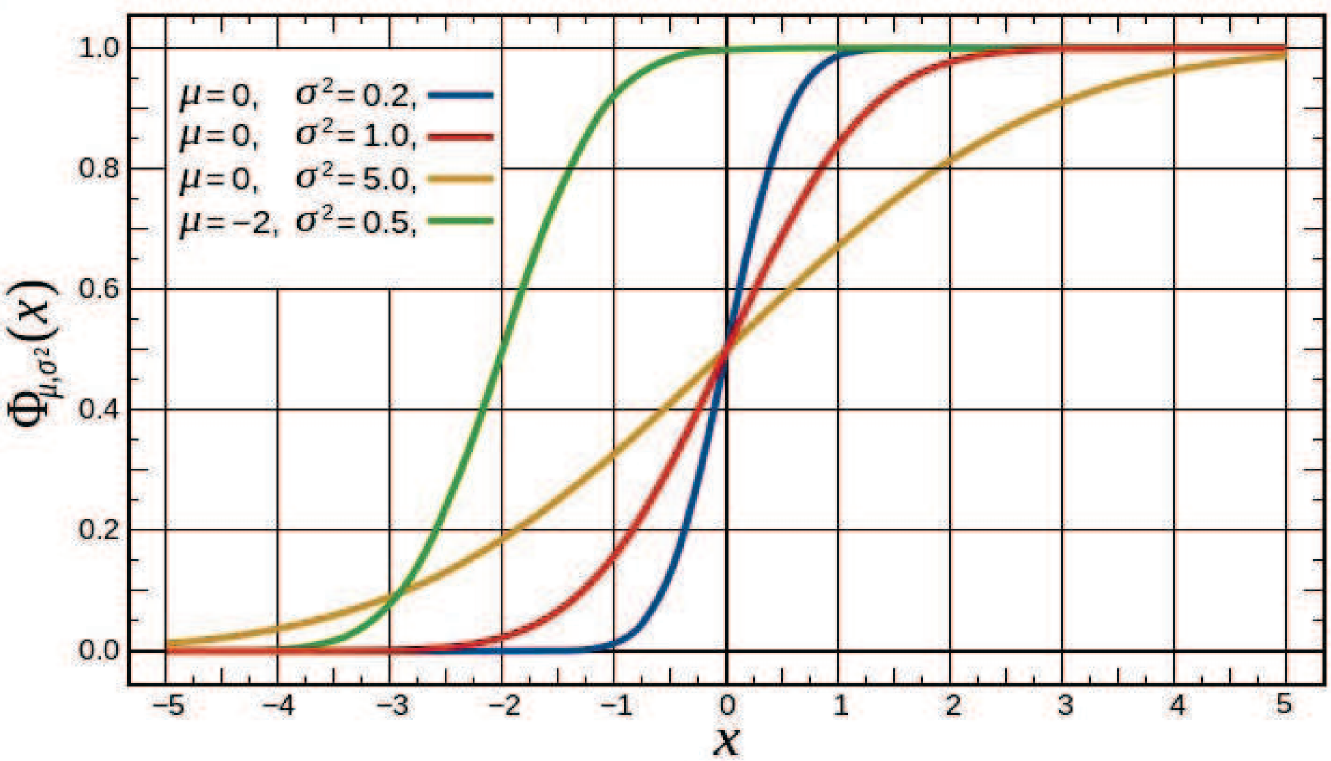

Figure 2: CDF of normal distribution.

2.1 Random Variables and Distribution Functions : CDF

- **Properties of $F_X(\cdot)$:** Suppose $F_X(\cdot)$ is the CDF of some random variable X. Then

  (1) $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to+\infty} F_X(x) = 1$.

  (2) $F_X(x)$ is non-decreasing, *i.e.*, for any $x_1 < x_2$, $F_X(x_1) \leq F_X(x_2)$.

  (3) $F_X(x)$ is right-continuous, *i.e.*, for all x and $\delta > 0$,

$$lim_{\delta\to 0^+}[F_X(x+\delta) - F_X(x)] = 0.$$

- **Theorem.** Let $a < b$. Then

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

- **Proof:** $\{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\}$ and the events $\{X \leq a\}$ and $\{a < X \leq b\}$ are disjoint, so we have

$$P(X \leq b) = P(X \leq a) + P(a < X \leq b).$$

- **Remarks:**

  – By definition, for any $b \in \mathbb{R}$,

$$\begin{aligned} P(X \geq b) &= P(X > b) + P(X = b) \\ &= 1 - F_X(b) + P(X = b). \end{aligned}$$

  – If a function $F(x)$ satisfies Property (1),(2),(3), then there exists a random variable X such that $P(X \leq x) = F(x)$.

  – Suppose $F_1(x)$ and $F_2(x)$ are two cumulative distribution functions, then for $0 < p < 1$,

$$F(x) = pF_1(x) + (1 - p)F_2(x)$$

  is also a cdf.

2.1 Random Variables and Distribution Functions : CDF

- **Theorem.** $P(X > b) = 1 - F_X(b)$.

- **Proof:** Let $A = \{X \leq a\}$. The the result follows from $P(A^c) = 1 - P(A)$ and the definition of the CDF of $F_X(x)$.

- **Definition: Identical Distributions.** Two random variables X and Y are identically distributed if for every set in $\mathbb{B}_1$, where $\mathbb{B}_1$ is the smallest σ-field containing all the intervals of real numbers of the form (a, b), $[a, b)$, $(a, b]$, and $[a, b]$, one has

$$P(X \in A) = P(Y \in A).$$

- **Question:** Does the identical distribution imply $X = Y$?

2.1 Random Variables and Distribution Functions : CDF

- **Example.** Suppose a penny and a nickel are each tossed n times, and consider the following two definitions of X and Y respectively:

  (1) X is the number of heads obtain with the penny, and Y is the number of heads obtained with the nickel.

  (2) Both X and Y are the number of heads obtained with the penny.

- In both cases, X and Y have the identical distribution. However, X and Y are independent in case (1) while $X = Y$ in case (2). Identical distribution does not imply $X = Y$, although $X = Y$ implies that X and Y have the same distribution.

- **Theorem.** Two random variables X and Y are identically distributed if and only if

$$F_X(x) = F_Y(x) \text{ for all } -\infty < x < \infty.$$

2.1 Random Variables and Distribution Functions : CDF

- **Example: First Order Stochastic Dominance.** If two distributions A and B, characterized respectively by CDFs F_A and F_B satisfy $F_A(x) \geq F_B(x)$ for all x, then we say that the distribution in B has first order stochastic dominance over distribution A. (See Figure 3.2 in lecture notes).

$$F(x) = \begin{cases} 1 - e^{-x}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0, \end{cases}$$

$$G(x) = \begin{cases} 1 - e^{-2x}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$$

Then $F(x) \leq G(x)$ for all $x \in (-\infty, \infty)$ and $F(\cdot)$ dominates G in first order.

- The first order stochastic dominance is widely used in decision analysis, welfare economics and so on. E.g., the analysis of income distribution.

2.2 Discrete Random Variable

- **Definition: Discrete Random Variable.** If a random variable X can only take a **countable** number of values, then X is called a *discrete random variable* (DRV).

- **Definition: Support of DRV.** The collection of the points on the real line $\mathbb{R}$ at which a DRV X has a positive probability is called the *support* of X, denoted as

$$\text{Support}(X) = \{x \in \mathbb{R} : f_X(x) > 0\}.$$

- **Remark:** Support of DRV is a countable set. Support$(X) = \Omega$.

2.2 Discrete Random Variable

- **Definition: Probability Mass Function (PMF).** Probability mass function of a DRV X is defined as

$$f_X(x) = P(X = x),\ x \in \text{Support}(X).$$

- **Theorem: Properties of PMF.**

(1) $0 \leq f_X(x) \leq 1$ for all $x \in \mathbb{R}$.

(2) $\sum_{x \in \Omega} f_X(x) = 1$.

- **Theorem.** Suppose $f_X(x)$ is the PMF of a DRV X. Then the CDF of a DRV X is

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= \sum_{y \leq x,\ y \in \text{Support}(X)} f_X(y),\ \text{for any } x \in \mathbb{R}. \end{aligned}$$

- **Example: Uniform Distribution.** A DRV X follows uniform distribution $U(N)$ if its PMF

$$f_X(x) = \begin{cases} 1/N & \text{if } x = 1, 2, \cdots, N, \\ 0 & \text{otherwise.} \end{cases}$$

- CDF of uniform distributed DRV X:
  - For $x < 1$, $F_X(x) = 0$.
  - For $i \le x < i + 1$, $i = 1, 2, \cdots, N - 1$, $F_X(x) = P(X \le x) = i/N$.
  - For $x \ge N$, $F_X(x) = 1$.
- See Figure 3.5 for $N = 6$.

2.2 Discrete Random Variable

- **Theorem.** Suppose X is a DRV with CDF $F_X(x)$, and its support contains a sequence of points $\{x_1 \leq x_2 \leq \cdots\}$. Then its PMF

$$f_X(x_i) = \begin{cases} F_X(x_i) & i = 1, \\ F_X(x_i) - F_X(x_{i-1}) & i > 1. \end{cases}$$

- **Remarks:**
  - $f_X(x)$ and $F_X(x)$ are equivalent ways to describe the DRV X.
  - $F_X(x_i) = F_X(x_{i-1}) + f_X(x_i)$. The CDF $F_X(x)$ always have jumps at points with strictly positive probabilities.

- **Example: Bernoulli Distribution.** A DRV X is called a Bernoulli(p) ($0 < p < 1$) random variable if its PMF

$$f_X(x) = \begin{cases} p & \text{if } x = 1, \\ 1 - p & \text{if } x = 0. \end{cases}$$

- **Example: Binomial Distribution.** A DRV X is called a Binomial(n, p) ($n \geq 0$ and $0 < p < 1$) if its PMF is

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \, x = 0, 1, \cdots, n$$

Remark: A person throws a coin n times independently. Each time the head has probability p and the tail has probability $q = 1 - p$. The number of heads is a random variable following Binomial(n, p) distribution.

- **Example: Poisson Distribution.** A DRV X follows a Poisson(λ) ($\lambda > 0$) distribution if its PMF

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \; x = 0, 1, 2, \cdots$$

- **Remarks:**
  - Support of Poisson distribution is a infinite countable set.
  - $\sum_{x=0}^{\infty} f_X(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1.$
  - In a Poisson process with intensity λ, the total number of occurrences over $(0, t]$ follows a Poisson(λt) distribution.
  - Poisson distribution can be used to describe the number of jumps in financial markets in a certain period.

- **Example: Negative Binomial Distribution.** The probability distribution of the number of trials required to obtain a given number of successes for binomial distribution is called the negative binomial distribution, denoted as $NB(n, p)$. That is, in a sequence of independent Bernoulli(p) trials, X denotes the number of trials such that the X-th trial the r-th success occurs, where r is a fixed integer. Then the PMF of X is

$$f_X(x) = \begin{pmatrix} x-1 \\ r-1 \end{pmatrix} p^r (1-p)^{x-r}, \ x = r, r+1, \ldots .$$

- **Example: Geometric Distribution.** The geometric distribution is the probability distribution of the number of Bernoulli trials required to obtain the first success. It is a special case of the negative binomial distribution with $r = 1$.

$$f_X(x) = p(1-p)^{x-1}, \ x = 1, 2, \ldots .$$

- **Remark:** The geometric distribution has "memoryless" property. For integers $s > t$, $P(X > s|X > t) = P(X > s - t)$.

$$
\begin{aligned}
P(X > s|X > t) &= \frac{P(X > s, X > t)}{P(X > t)} = \frac{P(X > s)}{P(X > t)} \\
&= \frac{1 - P(X \leq s)}{1 - P(X \leq t)} \\
&= \frac{1 - \sum_{x=1}^{s} p(1-p)^{x-1}}{1 - \sum_{x=1}^{t} p(1-p)^{x-1}} \\
&= \frac{(1-p)^s}{(1-p)^t} = (1-p)^{s-t} \\
&= P(X > s - t).
\end{aligned}
$$

2.3 Continuous Random Variables

- **Definition: Continuous Random Variables (CRV).** A random variable X is called *continuous* if its distribution function $F_X(x)$ is continuous for all x. In contrast, a random variable X is discrete if $F_X(x)$ is a step function of x.

- **Question:** Can we define a PMF $f_X(x)$ for a CRV X?

  For any constant $\varepsilon > 0$, $\{X = x\} \subset \{x - \varepsilon/2 < X \leq x + \varepsilon/2\}$,

$$\begin{aligned} 0 \ &\leq \ P(X = x) \leq P(x - \varepsilon/2 < X \leq x + \varepsilon/2) \\ &= \ F_X(x + \varepsilon/2) - F_X(x - \varepsilon/2) \to 0, \varepsilon \to 0. \end{aligned}$$

- So $P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b)$.

2.3 Continuous Random Variables

- **Definition: Absolute Continuity (AC).** A function $F : \mathbb{R} \to \mathbb{R}$ is called *absolutely continuous* with respect to Lebesgue measure if $F(x)$ is continuous on $\mathbb{R}$ and is differentiable almost everywhere (i.e. for almost all x).

- **Definition: Probability Density Function (PDF).** Suppose the distribution function $F_X(x)$ of a CRV X is absolutely continuous. Then there exists a function $f_X(\cdot)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy, \quad \text{for all } x \in \mathbb{R}.$$

The function $f_X(x) : \mathbb{R} \to \mathbb{R}$ is called a *probability density function*of X.

- **Remarks:**
  - For those x's where the derivative $F'_X(x)$ exists, the probability density function is
$$f_X(x) = \frac{dF_X(x)}{dx} = F'_X(x).$$
  - Because $f_X(x)$ is a slope of $F_X(x)$, it can take values greater than 1.

- **Question:** What is the interpretation of the pdf $f_X(x)$?

$$\begin{aligned} & P(x - \varepsilon/2 < X \leq x + \varepsilon/2) \\ = \; & F_X(x + \varepsilon/2) - F_X(x - \varepsilon/2) \\ = \; & \int_{x-\varepsilon/2}^{x+\varepsilon/2} f_X(x)dy = f_X(\bar{x})\varepsilon, \bar{x} \in (x - \varepsilon/2, x + \varepsilon/2] \end{aligned}$$

- **Remarks:**
  - An absolutely continuous CDF $F_X(x)$ is continuous for all $x \in \mathbb{R}$.
  - For some continuous CDF, absolute continuity may not hold.
  - The probability density functions of a continuous random variable can be different on a set of Lebesgue measure 0. For example, the pdf

$$f_X(x) = \begin{cases} e^{-x} & x > 0, \\ 0 & \text{elsewhere}, \end{cases}$$

can also be written as

$$f_X(x) = \begin{cases} e^{-x} & x \geq 0, \\ 0 & \text{elsewhere}. \end{cases}$$

  - Usually, we want $f_X(x)$ to be as smooth as possible.

2.3 Continuous Random Variable

- **Theorem: Properties of PDF.** A function $f(x)$ is a PDF of a CRV X iff (if and only if)

(1) $f(x) \geq 0$, for all x, and

(2) $\int_{-\infty}^{\infty} f(x)dx = 1$.

- **Example.** For any nonnegative function $g(x)$ with a finite integral, i.e., $0 < \int_{-\infty}^{\infty} g(x)dx < \infty$, $f(x) = g(x)/\int_{-\infty}^{\infty} g(y)dy$ is a PDF. $\int_{-\infty}^{\infty} g(y)dy$ is called the *normalizing constant*.

- **Definition: Support.** The support of a CRV X is defined as

$$\text{Support}(X) = \{x \in \mathbb{R} : f_X(x) > 0\},$$

where $f_X(x)$ is the PDF of X.

- **Example: Uniform Distribution.** A CRV X follows a uniform distribution on $[a, b]$ if its PDF

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

- **Example: Cauchy Distribution.** A CRV X follows a Cauchy(μ, σ) distribution if its PDF

$$f_X(x) = \frac{1}{\pi\sigma} \frac{1}{1 + (\frac{x-\mu}{\sigma})^2}, \text{ for } -\infty < x < \infty, \text{ where } \sigma > 0.$$

$\mu = 0, \sigma = 1$ is the standard Cauchy distribution Cauchy(0,1).

- **Example: Gamma Distribution.** A CRV X follows a $Gamma(\alpha, \beta)$ distribution if

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} & 0 < x < \infty, \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha, \beta > 0$, $\Gamma(\alpha)$ is the Gamma function $\int_0^\infty t^{\alpha-1} e^{-t} dt$.

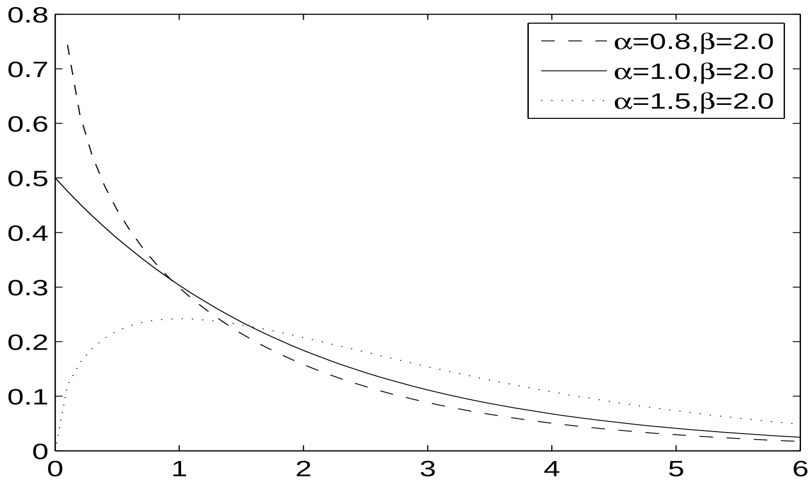

Figure 3: PDF of Gamma distribution.

- **Example: Exponential Distribution.** A CRV X follows Exponential(β) distribution if

$$f_X(x) = \begin{cases} \frac{1}{\beta}e^{-x/\beta} & 0 < x < \infty, \\ 0 & \text{otherwise}, \end{cases}$$

where $\beta > 0$.

- **Remarks:**

  – Exponential(β)=Gamma(1,β).

  – Exponential Distribution is popular in modelling duration between financial events or economic events because of its "memoryless" property

$$P(X - t > s \mid X - t > 0) = P(X > s \mid X > 0),$$

where $s > 0$, $t > 0$.

- **Example: Normal Distribution.** A normally distributed random variable, $X \sim N(\mu, \sigma^2)$, has the pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\{-\frac{(x-\mu)^2}{2\sigma^2}\},$$

where $-\infty < \mu < \infty$, $\sigma > 0$. X follows *standard normal* distribution if $\mu = 0$ and $\sigma^2 = 1$.

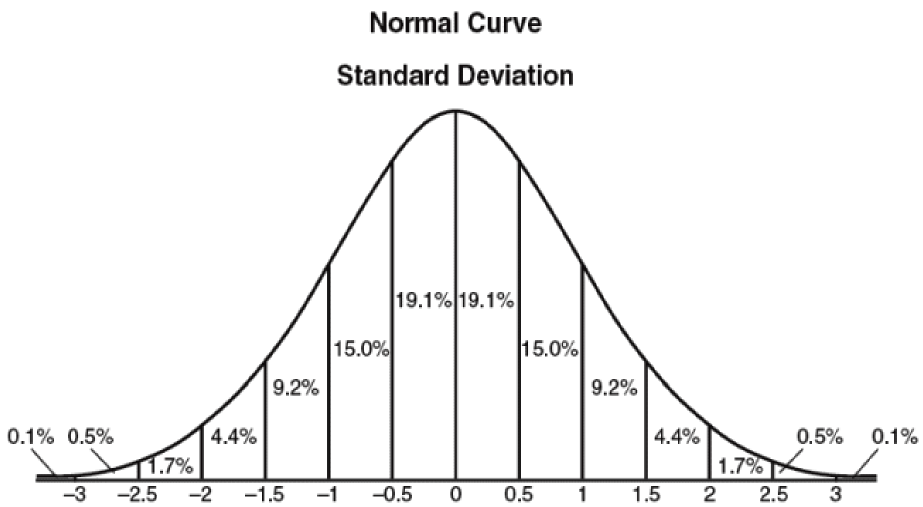

Figure 4: PDF of standard normal distribution.

2.3 Continuous Random Variable

- **Example: Log-normal Distribution.** X follows a log-normal(μ, σ^2) distribution if its pdf

$$f_X(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}\, exp\{-\frac{1}{2\sigma^2}(\log x - \mu)^2\} & 0 < x < \infty, \\ 0 & \text{otherwise,} \end{cases}$$

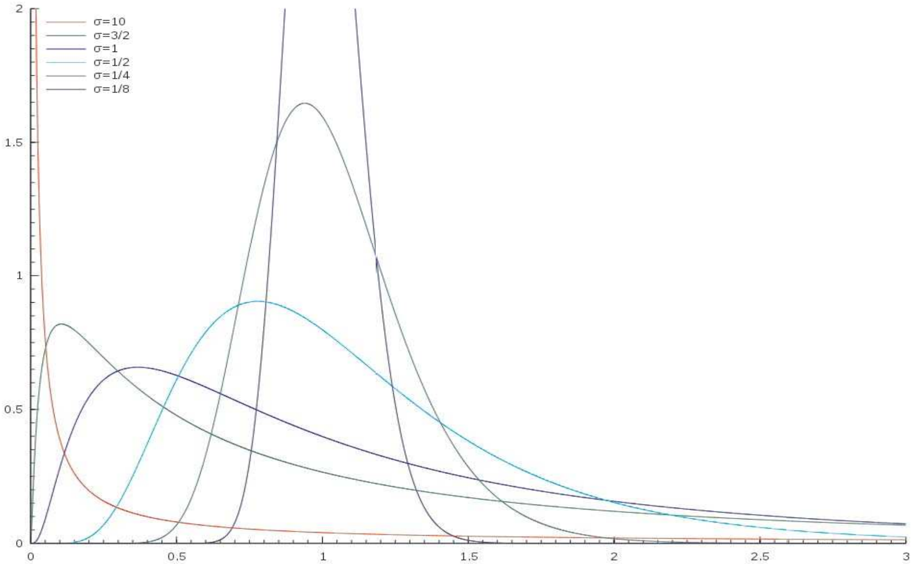

Figure 5: PDF of log-normal distribution.

- **Remarks:**
  - If X follows a log-normal(μ, σ) distribution, then $\log X \sim N(\mu, \sigma^2)$.
  - In finance, we often assume price P_t follows stochastic diffusion equation

$$dP_t = \mu_t P_t dt + \sigma P_t dW_t,$$

where W_t is a Brownian motion, μ and σ are constants. Using Ito's lemma,

$$d \log P_t = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW_t.$$

Here $\log P_t$ is normal, and P_t is log-normal.

- **Example: Chi-Square Distribution.** A nonnegative CRV X follows a Chi-Square distribution with p degrees of freedom, noted as χ_p^2, if its PDF

$$f_X(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{p/2-1} e^{-x/2}, \; x > 0.$$

The χ_p^2 distribution is a special case of the Gamma(α, β) distribution with $\alpha = p/2$, and $\beta = 2$. And it is equivalent to that of the sum of p squared independent $N(0,1)$ random variables.

- **Example: Double Exponential Distribution.** A continuous random variable X follows a double exponential distribution if its PDF

$$f_X(x) = \frac{1}{2\beta} \exp\big(- \frac{|x-\alpha|}{\beta} \big), \; -\infty < x < \infty, \text{where } \beta > 0.$$

- **Example: Beta Distribution.** X follows a Beta(α,β) distribution if

$$f_X(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} & 0 < x < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha, \beta > 0$, and $B(\alpha, \beta)$ is the Beta function $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$.

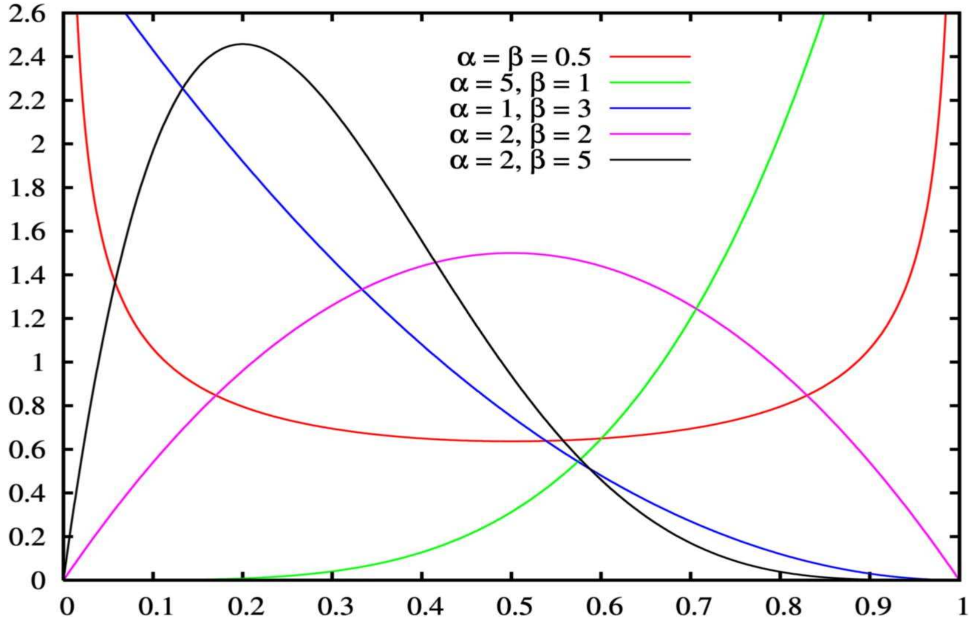

Figure 6: PDF of Beta distribution.

2.3 Continuous Random Variable

- **Definition: Mixed Distribution of Discrete and Continuous Components.** A random variable X is said to follow a mixed distribution if its CDF is discontinuous at each point having a nonzero probability and continuous elsewhere.

- **Lebesgue's Decomposition Theorem:** Any CDF $F_X(x)$ may be written in the form

$$F_X(x) = a_1F_1(x) + a_2F_2(x) + a_3F_3(x),$$

where $a_i \geq 0$, $i = 1, 2, 3$, $a_1 + a_2 + a_3 = 1$, $F_1(x)$ is absolutely continuous, $F_2(x)$ is a step function with a finite or countably infinite number of jumps, $F_3(x)$ is a singular CDF. That is, it is continuous with zero derivative almost everywhere.

2.4 Functions of a Random Variable

- **Question:** Suppose $g : \mathbb{R} \to \mathbb{R}$ is a real-valued (Borel)-measurable function, then $Y = g(X)$ is also a random variable. What is the probability distribution of the new random variable Y?

- **Examples:**

  – Consumption function $Y = g(X)$, where X is income and Y is consumption.

  – Assume P_t is the stock price, $Y_t = \log P_t/P_{t-1}$ is approximately the relative price change.

2.4 Functions of a Random Variable : DRV

Discrete Case:

- If X is a **discrete** random variable, then the pmf $f_Y(y)$ of $Y = g(X)$ given the pmf $f_X(x)$ of X can be obtained by using

$$f_Y(y) = \sum_{x:g(x)=y} f_X(x),$$

where the summation is over all possible x's whose $g(x) = y$.

- **Example:** Suppose random variable X has the distribution

x	-2	-1	0	1	2
$f_X(x)$	0.2	0.1	0.1	0.3	0.3

Find CDF of function (transformation) $Y = X^2 + X$.

2.4 Functions of a Random Variable : CRV

Continuous Case:

- If X is a continuous random variable, the basic idea is first to find the distribution function of Y and then the probability density by differentiation.

- **General Method:**

  – Step 1: Find $F_Y(y)$

  $$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(\{x \in \Omega_X : g(x) \leq y\}).$$

  – Step 2: Let

  $$f_Y(y) \; = \; F'_Y(y).$$

  – Step 3: Check if $f_Y(y)$ is a PDF.

2.4 Functions of a Random Variable : CRV

- **Example:** Suppose a CRV X has a pdf

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

Find the pdf of the following Y.

(1) $Y = a + bX$, $b \neq 0$.

(2) $Y = X^2$.

(3) $Y = |X|$.

- **Solution:**

(2) Observe that $0 \le X^2 < 1/4$, then $F_Y(y) = 0$ if $y < 0$ and $F_Y(y) = 1$ if $y > 1/4$. For $y \in [0, 1/4]$,

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= 2\sqrt{y}. \end{aligned}$$

By differentiation,

$$f_Y(y) = \frac{1}{\sqrt{y}}.$$

Thus, we have

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{y}} & 0 < y < 1/4, \\ 0 & \text{otherwise.} \end{cases}$$

- **Example:** Random variable X follows *double exponential* (Laplace) distribution if

$$f_X(x) = \frac{1}{2}\alpha e^{-\alpha|x|},$$

where $\alpha > 0$. Find the pdf of the following Y.

(1) $Y = |X|$;

(2) $Y = X^2$.

- **Solution:**

(1) $F_Y(y) = 0$ for $y \leq 0$. For $y > 0$,

$$F_Y(y) = F_X(-y \leq X \leq y) = \int_{-y}^{y} f_X(x)dx$$

By differentiation,

$$f_Y(y) = \frac{1}{2}\alpha e^{-\alpha y} + \frac{1}{2}\alpha e^{-\alpha|-y|} = \alpha e^{-\alpha y},$$

for $y > 0$ and $f_Y(y) = 0$ for $y \leq 0$. It is an exponential$(1/\alpha)$.

- (2) $F_Y(y) = 0$ for $y \leq 0$. For $y > 0$,

$$F_Y(y) = F_X(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx$$

By differentiation,

$$\begin{aligned} f_Y(y) &= \frac{1}{2}\alpha e^{-\alpha\sqrt{y}} \times \frac{d(\sqrt{y})}{dy} - \frac{1}{2}\alpha e^{-\alpha|-\sqrt{y}|}\frac{d(-\sqrt{y})}{dy} \\ &= \frac{1}{2}\alpha e^{-\alpha\sqrt{y}} \times \frac{1}{2\sqrt{y}} + \frac{1}{2}\alpha e^{-\alpha|-\sqrt{y}|}\frac{1}{2\sqrt{y}} = \frac{\alpha}{2\sqrt{y}}e^{-\alpha y}, \end{aligned}$$

for $y > 0$ and $f_Y(y) = 0$ for $y \leq 0$.

- **Remark:** A Weibull distribution has a pdf $f_X(x) = \frac{\beta}{\delta}(\frac{x-\gamma}{\delta})^{\beta-1}\exp[-(\frac{x-\gamma}{\delta})^{\beta}]$, $x > \gamma$, and 0 otherwise. This distribution is widely used in survival analysis or duration analysis. Here $\beta = 1/2, \delta = \alpha^{-2}, \gamma = 0$.

- **Example.** Suppose $X \sim N(0,1)$. Find the PDF $f_Y(y)$ of $Y = X^2$.

- **Solution:** For $y \geq 0$, $P(Y \leq y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$. So,

$$\begin{aligned} f_Y(y) &= F_X'(\sqrt{y})\frac{1}{2\sqrt{y}} + F_X'(-\sqrt{y})\frac{1}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}}e^{-y/2}. \end{aligned}$$

  Y is a chi-square random variable with degree of freedom 1, χ_1^2.

- **Example.** Suppose $X \sim N(\mu, \sigma)$, then $Y = e^x$ follows *log-normal* distribution. Find the pdf of Y.

- **Solution:** For $y > 0$,

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{y}e^{-(\log y - \mu)^2/2\sigma^2}.$$

2.4 Functions of a Random Variable : C.R.V.

- **Example:** Suppose X has pdf $f_X(x) = \frac{3}{8}(x+1)^2$, $-1 < x < 1$, and

$$Y = \begin{cases} 1 - X^2 & \text{if } X \le 0, \\ 1 - X & \text{if } X > 0. \end{cases}$$

Find the pdf of Y.

- **Solution:**

  – Observe that $0 < y \le 1$.

  – For $0 < y \le 1$,

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(-1 < X \le -\sqrt{1-y}) + P(1 - y \le X < 1) \\ &= \int_{-1}^{-\sqrt{1-y}} f_X(x)dx + \int_{1-y}^{1} f_X(x)dx \end{aligned}$$

- – By differentiation,

$$\begin{aligned} f_Y(y) &= f_X(-\sqrt{1-y})\frac{d(-\sqrt{1-y})}{dy} - f_X(1-y)\frac{d(1-y)}{dy} \\ &= \frac{3}{8}(1-\sqrt{1-y})^2\frac{1}{2\sqrt{1-y}} + \frac{3}{8}(2-y)^2, \end{aligned}$$

  and $f_Y(y) = 0$ for $y \le 0$ or $y > 1$.

- **Theorem: Probability Integral Transform.** Suppose X has a continuous distribution $F_X(x)$ which is strictly monotonically increasing. Define $Y = F_X(X)$, that is

$$Y = \int_{-\infty}^{X} f_X(x)dx.$$

  Then Y follows a uniform distribution on $[0, 1]$.

- **Proof:** Because $F_X(x)$ is continuous and strictly monotonically increasing, $F_X(\cdot)$ forms a 1-1 correspondence between $\mathbb{R}$ and interval $[0,1]$. Inverse function $F_X^{-1}(\cdot)$ exists and $F_X(F_X^{-1}(y)) = y$, for any $y \in \mathbb{R}$. For $0 \leq y \leq 1$,

$$
\begin{aligned}
F_Y(y) &= P(F_X(x) \leq y) \\
&= P(x \leq F_X^{-1}(y)) \\
&= F_X(F_X^{-1}(y)) \\
&= y.
\end{aligned}
$$

It follows that the pdf

$$
f_Y(y) = \begin{cases} 1 & 0 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}
$$

Y follows Uniform$[0,1]$ distribution.

2.4 Functions of a Random Variable : C.R.V.

- **Example:** How to generate random sample $X \sim$ exponential(1)?

- **Solution:**

  – The cdf of X is

$$F_X(x) = \begin{cases} 1 - e^{-x} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

  Let $Y = F_X(X) = 1 - e^{-X}$, then $Y \sim U(0, 1)$.

  – Now generate $Y \sim U[0, 1]$ and let

$$X = -\log(1 - Y) \sim \text{exponential}(1).$$

- **Remark.** The result that $F_X(X) \sim U[0, 1]$ provides a basis for goodness-of-tests of distributional models. And it is the basic idea behind the Kolmogorov-Smirnov test for a hypothesized distribution model.

2.4 Functions of a Random Variable : C.R.V.

- **Alternative Method: The Transformation Approach.**

- **Theorem: Univariate Transformation.** Let X be a CRV with PDF $f_X(x)$ and let function $g : \mathbb{R} \to \mathbb{R}$ be **strictly monotone** and differentiable over the support of X. Then the PDF of the random variable $Y = g(X)$ is

$$f_Y(y) = \frac{1}{|g'(x)|} f_X(x) \,|_{x=g^{-1}(y)},$$

for any y in the support of Y; where x is the **unique** number in the support of X such that $g(x) = y$.

2.4 Functions of a Random Variable : C.R.V.

- **Proof:** When $g(x)$ is strictly increasing, $\exists$ a unique strictly increasing inverse function $g^{-1}(y)$ s.t. $g^{-1}[g(x)] = x$. For y in the support of Y,

$$F_Y(y) = P(Y \le y) = P[g(X) \le y] = P[X \le g^{-1}(y)] = F_X[g^{-1}(y)].$$

By the chain rule of differentiation, we obtain

$$f_Y(y) = F'_Y(y) = F'_X[g^{-1}(y)]\frac{d}{dy}g^{-1}(y) = f_X(x)\frac{1}{g'(x)},$$

where $x = g^{-1}(y)$, and $\frac{d}{dy}g^{-1}(y) = \frac{1}{g'(x)}$, $g^{-1}(y) = x$.

When $g(X)$ is monotonically decreasing. Similarly, we can obtain $f_Y(y) = -f_X(x)\frac{1}{g'(x)}$, which finishes the proof.

2.4 Functions of a Random Variable : C.R.V.

- **Remark.** Before applying this transformation theorem, it is very important to check whether the function g is strictly monotonic. It can not be directly applied to non-monotonic functions. However, if $Y = g(X)$ is strictly monotonic over several regions of the real line, we can extend the above univariate transformation theorem to cover such more general cases.

- **Theorem.** Suppose $g(x) = g_i(x)$ for all $x \in A_i$, where $i = 1, 2, ..., k$, where for each i, $g_i(x)$ is strictly monotonic (strictly increasing or decreasing) and diffentiable on region A_i, and the regions $\{A_i\}$ are disjoint and $\cup_{i=1}^{k} A_i = R$. Then the PDF of $Y = g(X)$ is given by

$$f_Y(y) = \sum_{i=1}^{k} f_X[g_i^{-1}(y)] \frac{1}{| \, g_i'[g_i^{-1}(y)]} \, |$$

for all y in the support of Y.