COM 5115: Stochastic Processes for Networking

Prof. Shun-Ren Yang
Department of Computer Science,
National Tsing Hua University, Taiwan

Outline

- Preliminaries
- Poisson Processes
- Renewal Processes
- Discrete-Time Markov Chains
- Continuous-Time Markov Chains

Preliminaries

- Applied Probability and Performance Modeling
  - Prototype
  - System Simulation
  - Probabilistic Model
- Introduction to Stochastic Processes
  - Random Variable (R.V.)
  - Stochastic Process
- Probability and Expectations
  - Expectation
  - Generating Functions for Discrete R.V.s
  - Laplace Transforms for Continuous R.V.s
  - Moment Generating Functions

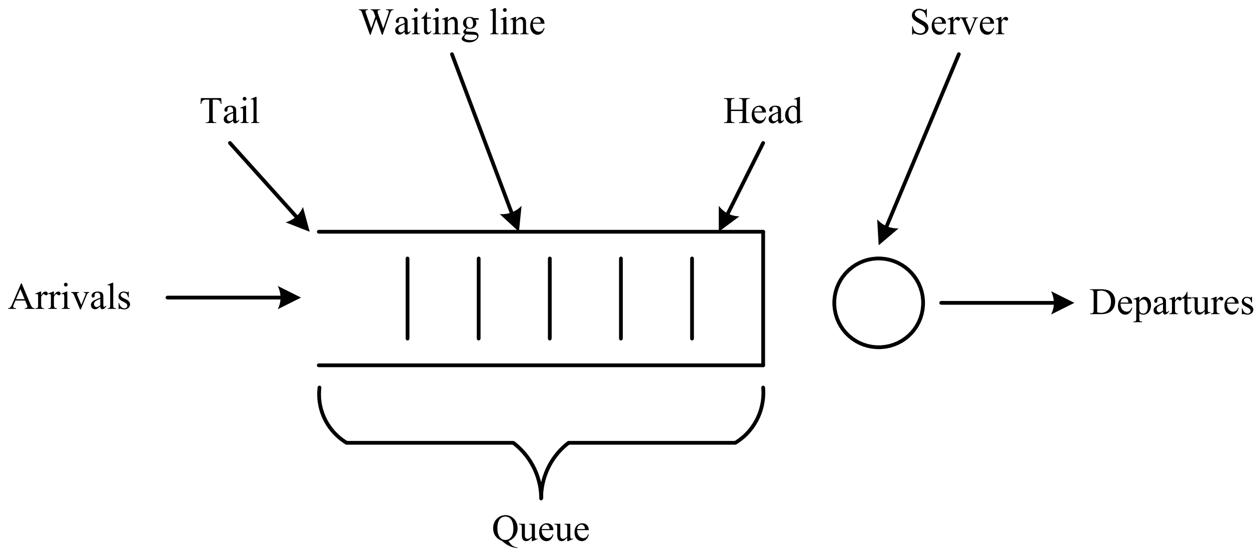
Preliminaries

- Probability Inequalities
  - Markov's Inequality (mean)
  - Chebyshev's Inequality (mean and variance)
  - Chernoff's Bound (moment generating function)
  - Jensen's Inequality
- Limit Theorems
  - Strong Law of Large Numbers
  - Weak Law of Large Numbers
  - Central Limit Theorem

Applied Probability and Performance Modeling

- Prototyping
  - complex and expensive
  - provides information on absolute performace measures but little on relative performance of different designs
- System Simulation
  - large amount of execution time
  - could provide both absolute and relative performance depending on the level of detail that is modeled
- Probabilistic Model
  - mathematically intractable or unsolvable
  - provide great insight into relative performance but, often, are not accurate representations of absolute performance

A Single Server Queue



- Arrivals: Poisson process, renewal process, etc.
- Queue length: Markov process, semi-Markov process, etc.
- ...

Random Variable

- A "*random variable*" is a real-valued function whose domain is a sample space.
- **Example.** Suppose that our experiment consists of tossing 3 fair coins. If we let $\tilde{y}$ denote the number of heads appearing, then $\tilde{y}$ is a random variable taking on one of the values 0, 1, 2, 3 with respective probabilities

$$\begin{aligned}
P\{\tilde{y}=0\} &= P\{(T,T,T)\} = \frac{1}{8} \\
P\{\tilde{y}=1\} &= P\{(T,T,H),(T,H,T),(H,T,T)\} = \frac{3}{8} \\
P\{\tilde{y}=2\} &= P\{(T,H,H),(H,T,H),(H,H,T)\} = \frac{3}{8} \\
P\{\tilde{y}=3\} &= P\{(H,H,H)\} = \frac{1}{8}
\end{aligned}$$

Random Variable

- A random variable $\tilde{x}$ is said to be "*discrete*" if it can take on only a finite number—or a countable infinity—of possible values x.
- A random variable $\tilde{x}$ is said to be "*continuous*" if there exists a nonnegative function f, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers

$$P\{\tilde{x} \in B\} = \int_B f(x)dx$$

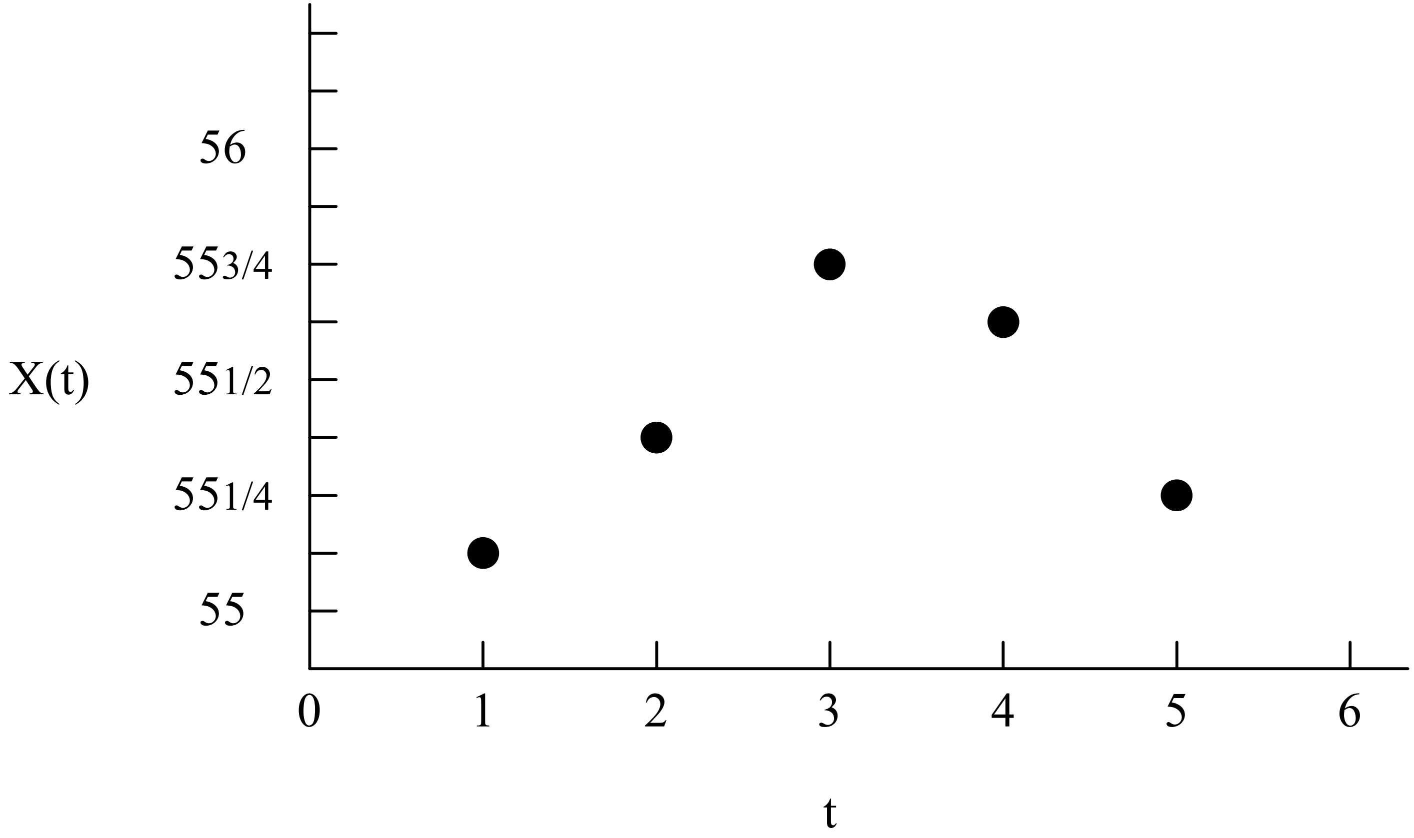
Stochastic Process

- A "*stochastic process*" $X = \{\tilde{x}(t), t \in T\}$ is a collection of random variables. That is, for each $t \in T$, $\tilde{x}(t)$ is a random variable.
- The index t is often interpreted as "*time*" and, as a result, we refer to $\tilde{x}(t)$ as the "*state*" of the process at time t.
- When the index set T of the process X is
  - a countable set $\rightarrow$ X is a *discrete-time* process
  - an interval of the real line $\rightarrow$ X is a *continuous-time* process
- When the state space S of the process X is
  - a countable set $\rightarrow$ X has a *discrete state space*
  - an interval of the real line $\rightarrow$ X has a *continuous state space*

Stochastic Process

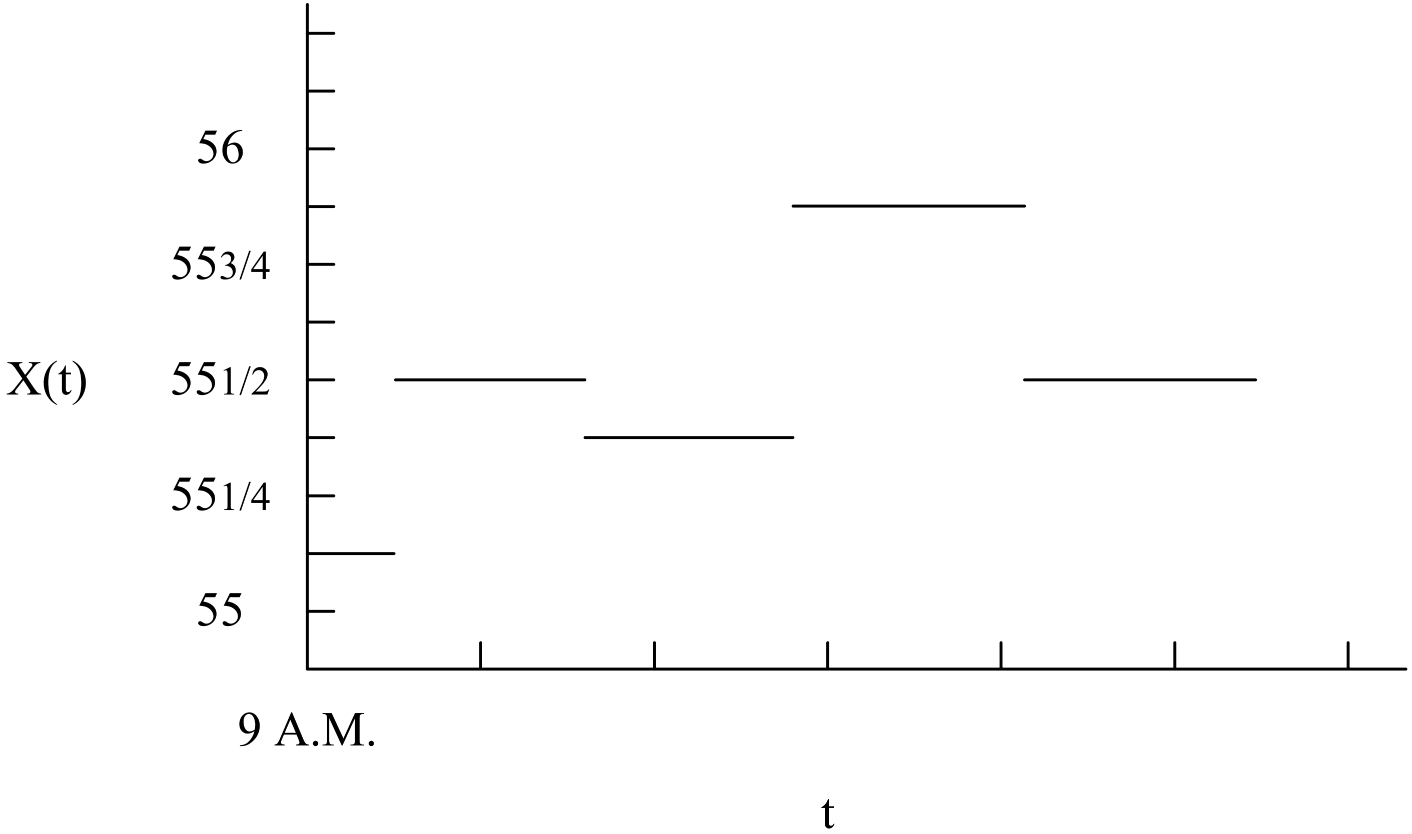
- Four types of stochastic processes
  - discrete time and discrete state space
  - continuous time and discrete state space
  - discrete time and continuous state space
  - continuous time and continuous state space

Discrete Time with Discrete State Space



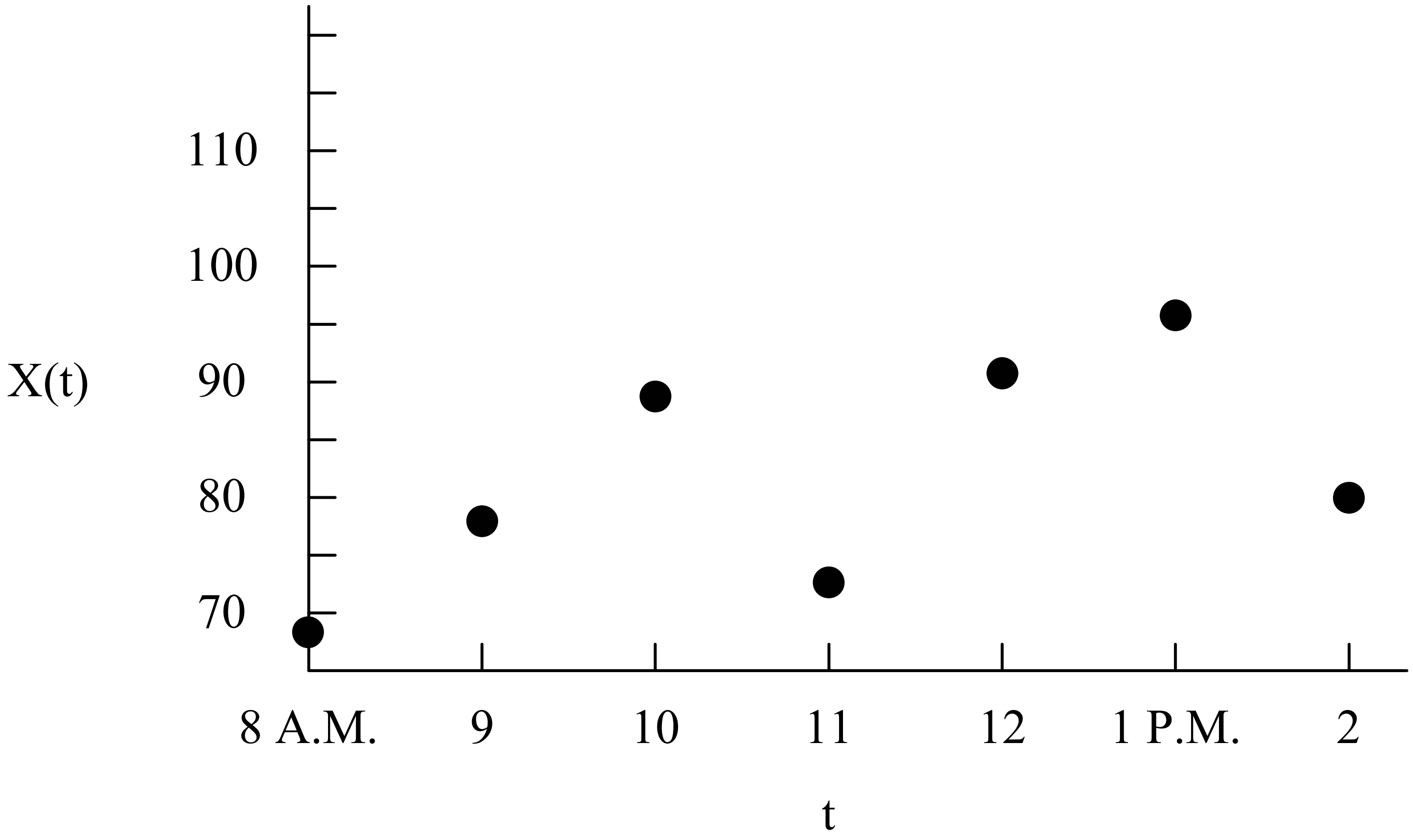
X(t) = closing price of an IBM stock on day t

Continuous Time with Discrete State Space



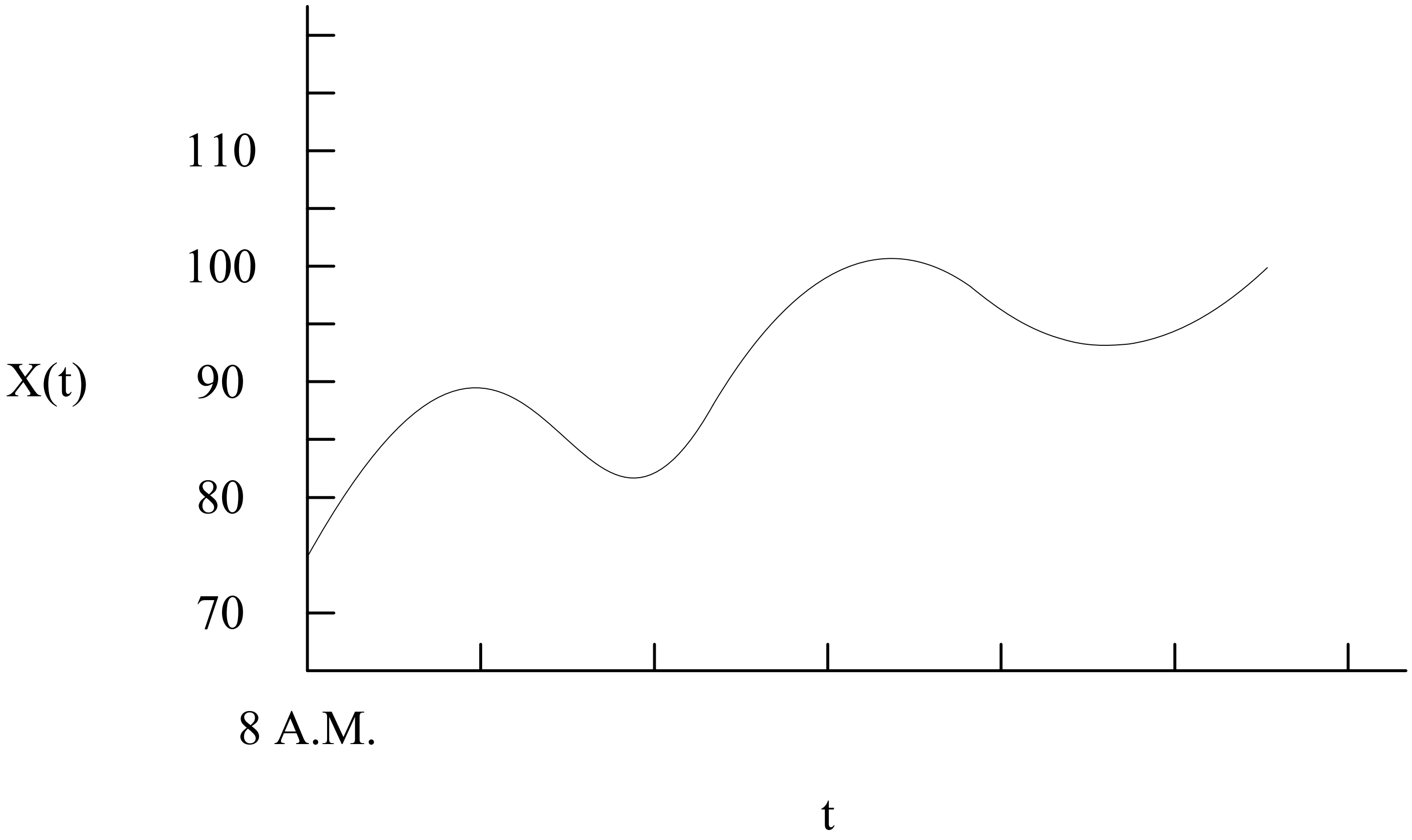
X(t) = price of an IBM stock at time t on a given day

Discrete Time with Continuous State Space



X(t) = temperature at the airport at time t

Continuous Time with Continuous State Space



X(t) = temperature at the airport at time t

Two Structural Properties of stochastic processes

a. Independent increment: if for all $t_0 < t_1 < t_2 < \ldots < t_n$ in the process $X = \{\tilde{x}(t), t \geq 0\}$, random variables $\tilde{x}(t_1) - \tilde{x}(t_0), \tilde{x}(t_2) - \tilde{x}(t_1), \ldots, \tilde{x}(t_n) - \tilde{x}(t_{n-1})$ are independent,
$\Rightarrow$ the magnitudes of state change over non-overlapping time intervals are mutually independent

b. Stationary increment: if the random variable $\tilde{x}(t+s) - \tilde{x}(t)$ has the same probability distribution for all t and any $s > 0$,
$\Rightarrow$ the probability distribution governing the magnitude of state change depends only on the difference in the lengths of the time indices and is independent of the time origin used for the indexing variable

$$\Downarrow$$

$$X = \{\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \ldots, \tilde{x}_\infty\}$$

limiting behavior of the stochastic process

Two Structural Properties of stochastic processes

<Homework>. Define stochastic processes that you think have the following properties:

- both independent and stationary increments,
- neither independent nor stationary increments,
- independent but not stationary increments, and
- stationary but not independent increments.

Expectations by Conditioning

Denote by $E[\tilde{x}|\tilde{y}]$ that function of the random variable $\tilde{y}$ whose value at $\tilde{y} = y$ is $E[\tilde{x}|\tilde{y} = y]$.

$\Rightarrow E[\tilde{x}] = E[E[\tilde{x}|\tilde{y}]]$

If $\tilde{y}$ is a discrete random variable, then

$$E[\tilde{x}] = \sum_y E[\tilde{x}|\tilde{y} = y]P\{\tilde{y} = y\}$$

If $\tilde{y}$ is continuous with density $f_{\tilde{y}}(y)$, then

$$E[\tilde{x}] = \int_{-\infty}^{\infty} E[\tilde{x}|\tilde{y} = y] f_{\tilde{y}}(y) dy$$

Expectations by Complementary Distribution

For any non-negative random variable $\tilde{x}$

$$E[\tilde{x}] = \sum_{k=0}^{\infty} p(\tilde{x} > k) \qquad \text{discrete}$$

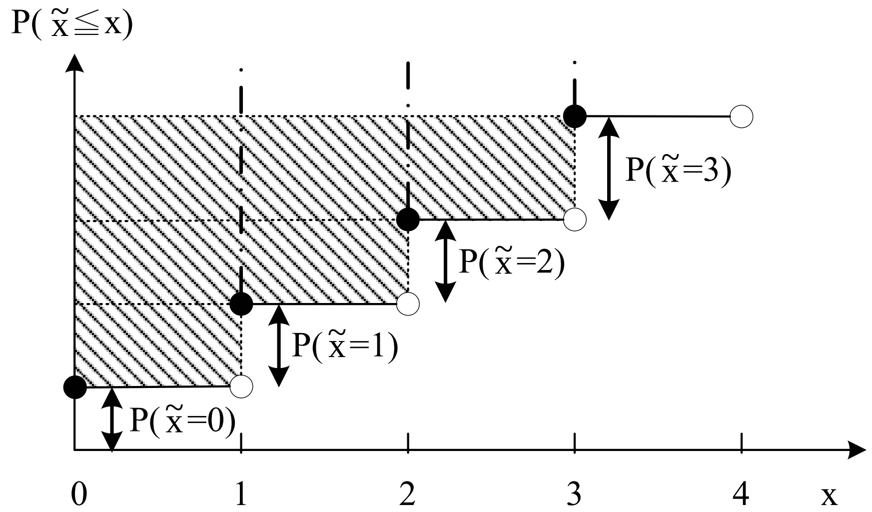
$$E[\tilde{x}] = \int_0^{\infty} [1 - F_{\tilde{x}}(x)]dx \qquad \text{continuous}$$

. . . .

Expectations by Complementary Distribution

<u>Discrete case:</u>

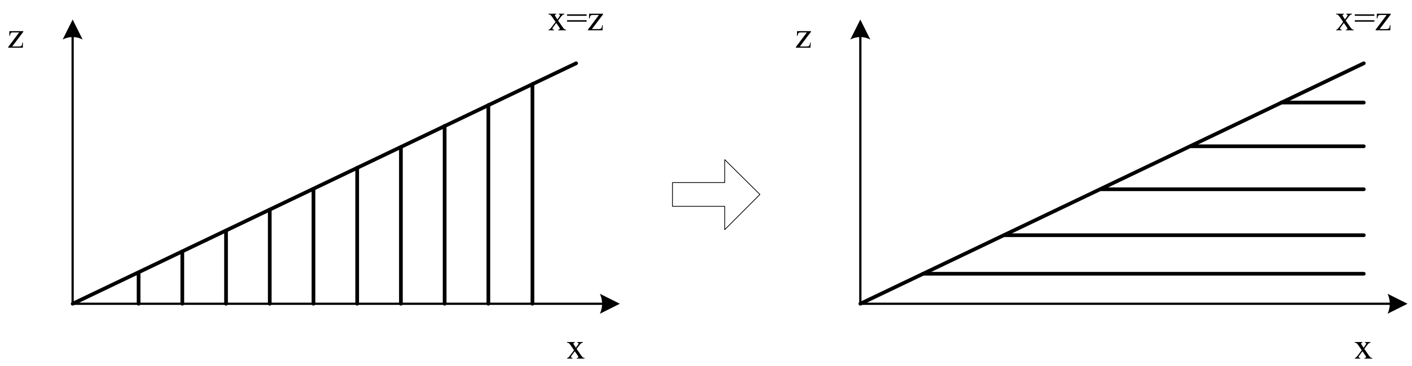
$$
\begin{aligned}
E[\tilde{x}] &= 0 \cdot P(\tilde{x}=0) + 1 \cdot P(\tilde{x}=1) + 2 \cdot P(\tilde{x}=2) + \ldots \ \text{(horizontal sum)} \\
&= [1 - P(\tilde{x} < 1)] + [1 - P(\tilde{x} < 2)] + \ldots \ \text{(vertical sum)} \\
&= P(\tilde{x} \geq 1) + P(\tilde{x} \geq 2) + \ldots \\
&= \sum_{k=1}^{\infty} P(\tilde{x} \geq k) \qquad (or \sum_{k=0}^{\infty} P(\tilde{x} > k))
\end{aligned}
$$



Expectations by Complementary Distribution

<u>Continuous case:</u>

$$
\begin{aligned}
E[\tilde{x}] &= \int_0^\infty x \cdot f_{\tilde{x}}(x)dx \\
&= \int_0^\infty \left(\int_0^x dz\right) \cdot f_{\tilde{x}}(x)dx \\
&= \int_0^\infty \left[\int_z^\infty f_{\tilde{x}}(x)dx\right] \cdot dz \\
&= \int_0^\infty [1 - F_{\tilde{x}}(z)]dz
\end{aligned}
$$



Compound Random Variable

$\tilde{S}_{\tilde{n}} = \tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 + \ldots + \tilde{x}_{\tilde{n}}$, where $\tilde{n} \geq 1$ and $\tilde{x}_i$ are i.i.d. random variables.

$\Rightarrow E[\tilde{S}_{\tilde{n}}] =?\ Var[\tilde{S}_{\tilde{n}}] =?$

. . . .

$$\begin{aligned} E[\tilde{S}_{\tilde{n}}] &= E[E[\tilde{S}_{\tilde{n}}|\tilde{n}]] \\ &= \sum_{n=1}^{\infty} E[\tilde{S}_{\tilde{n}}|\tilde{n}=n] \cdot P(\tilde{n}=n) \\ &= \sum_{n=1}^{\infty} E[\tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_n] \cdot P(\tilde{n}=n) \\ &= \sum_{n=1}^{\infty} n \cdot E[\tilde{x}_1] \cdot P(\tilde{n}=n) \\ &= E[\tilde{n}] \cdot E[\tilde{x}_1] \end{aligned}$$

Compound Random Variable

Since $Var[\tilde{x}] = E[Var[\tilde{x}|\tilde{y}]] + Var[E[\tilde{x}|\tilde{y}]]$, we have

$$\begin{aligned} Var[\tilde{S}_{\tilde{n}}] &= E[Var[\tilde{S}_{\tilde{n}}|\tilde{n}]] + Var[E[\tilde{S}_{\tilde{n}}|\tilde{n}]] \\ &= E[\tilde{n}Var[\tilde{x}_1]] + Var[\tilde{n}E[\tilde{x}_1]] \\ &= Var[\tilde{x}_1]E[\tilde{n}] + E^2[\tilde{x}_1]Var[\tilde{n}] \end{aligned}$$

Probability Generating Functions for Discrete R.V.s

- Define the *generating function* or *Z-transform* for a sequence of numbers $\{a_n\}$ as $a^g(z) = \sum_{n=0}^{\infty} a_n z^n$.

- Let $\tilde{x}$ denote a discrete random variable and $a_n = P[\tilde{x} = n]$. Then $P_{\tilde{x}}(z) = a^g(z) = \sum_{n=0}^{\infty} a_n z^n = E[z^{\tilde{x}}]$ is called the *probability generating function* for the random variable $\tilde{x}$.

- Define the kth derivative of $P_{\tilde{x}}(z)$ by

$$P_{\tilde{x}}^{(k)}(z) = \frac{d^k}{dz^k} P_{\tilde{x}}(z).$$

Then, we see that

$$P_{\tilde{x}}^{(1)}(z) = \sum_{n=0}^{\infty} n a_n z^{n-1} \;\rightarrow\; P_{\tilde{x}}^{(1)}(1) = E[\tilde{x}]$$

Probability Generating Functions for Discrete R.V.s

and

$$P_{\tilde{x}}^{(2)}(z) = \sum_{n=1}^{\infty} n(n-1)a_n z^{n-2} \ \rightarrow \ P_{\tilde{x}}^{(2)}(1) = E[\tilde{x}^2] - E[\tilde{x}]$$

- See Table 1.1 [Kao] for the properties of generating functions.
- <Homework>. Derive the probability generating functions for "Binomial", "Poisson", "Geometric" and "Negative Binomial" random variables. Then, derive the expected value and variance of each random variable via the probability generating function.

Laplace Transforms for Continuous R.V.s

- Let f be any real-valued function defined on $[0, \infty)$. The Laplace transform of f is defined as

$$F^*(s) = \int_0^\infty e^{-st} f(t)dt.$$

- When f is a probability density of a nonnegative continuous random variable $\tilde{x}$, we have

$$F^*_{\tilde{x}}(s) = E[e^{-s\tilde{x}}]$$

- Define the nth derivative of the Laplace transform $F^*_{\tilde{x}}(s)$ with respect to s by

$$F^{*(n)}_{\tilde{x}}(s) = \frac{d^n}{ds^n} F^*_{\tilde{x}}(s) \rightarrow F^{*(n)}_{\tilde{x}}(s) = (-1)^n E[\tilde{x}^n e^{-s\tilde{x}}].$$

Then, we see that

$$E[\tilde{x}^n] = (-1)^n F^{*(n)}_{\tilde{x}}(0)$$

Laplace Transforms for Continuous R.V.s

- See Table 1.2 [Kao] for the properties of Laplace Transforms.
- <Homework>. Derive the Laplace transforms for "Uniform", "Exponential", and "Erlang" random variables. Then, derive the expected value and variance of each random variable via the Laplace transform.

Moment Generating Functions

- The moment generating function $M_{\tilde{x}}(\theta)$ of the random variable $\tilde{x}$ is defined for all values θ by

$$\begin{aligned} M_{\tilde{x}}(\theta) &= E[e^{\theta\tilde{x}}] \\ &= \begin{cases} \sum_x e^{\theta x} p(x), & \text{if } \tilde{x} \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{\theta x} f(x) dx, & \text{if } \tilde{x} \text{ is continuous} \end{cases} \end{aligned}$$

- The nth derivative of $M_{\tilde{x}}(\theta)$ evaluated at $\theta = 0$ equals the nth moment of $\tilde{x}$, $E[\tilde{x}^n]$, that is,

$$M_{\tilde{x}}^{(n)}(0) = E[\tilde{x}^n], \quad n \geq 1$$

Markov's Inequality

- Let h be a nonnegative and nondecreasing function and let $\tilde{x}$ be a random variable. If the expectation of $h(\tilde{x})$ exists then it is given by

$$E[h(\tilde{x})] = \int_{-\infty}^{\infty} h(z) f_{\tilde{x}}(z) dz. \tag{1}$$

- By assumptions on h it easily follows that

$$\int_{-\infty}^{\infty} h(z) f_{\tilde{x}}(z) dz \geq \int_{t}^{\infty} h(z) f_{\tilde{x}}(z) dz \geq h(t) \int_{t}^{\infty} f_{\tilde{x}}(z) dz. \tag{2}$$

- Combining (1) and (2) yields Markov's inequality:

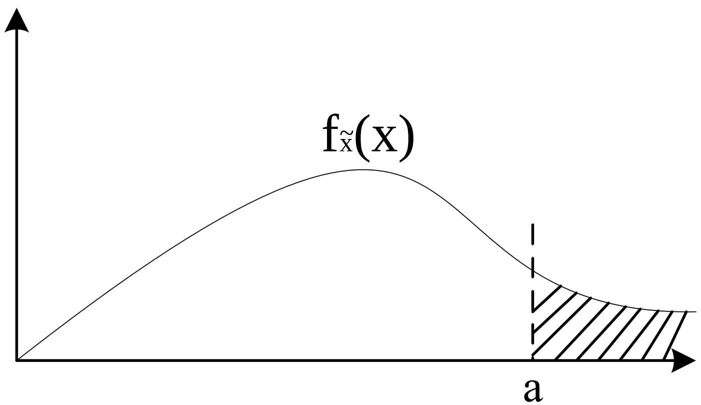
$$P[\tilde{x} \geq t] \leq \frac{E[h(\tilde{x})]}{h(t)}, \quad \textbf{Markov's Inequality.}$$

- When $h(x) = x$ and $\tilde{x}$ is nonnegative then we have

$$P[\tilde{x} \geq t] \leq \frac{E[\tilde{x}]}{t}, t > 0 \quad \textbf{Simple Markov's Inequality.}$$

Markov's Inequality

- The simple Markov's inequality is a *first-order* inequality since only knowledge of $E[\tilde{x}]$ is required.
- The simple Markov's inequality is quite weak but can be used to quickly check statements made about *the tail of a distribution* of a random variable when the expectation is known.



- **Example.** If the expected response time of a computer system is 1 second, then the simple Markov's inequality shows that $P[\tilde{x} \geq 10] \leq .1$ and thus at most 10% of the response times in the system can be greater than 10 seconds.

Chebyshev's Inequality – second-order bound

If $\tilde{x}$ is a random variable with mean μ and variance σ^2, $k > 0$, then

$$P(|\tilde{x} - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

$\underline{Proof:}$

Since $(\tilde{x} - \mu)^2$ is a non-negative random variable, applying Markov's inequality yields

$$\begin{aligned} P((\tilde{x} - \mu)^2 \geq k^2) &\leq \frac{E[(\tilde{x} - \mu)^2]}{k^2} \\ P(|\tilde{x} - \mu| \geq k) &\leq \frac{\sigma^2}{k^2} \end{aligned}$$

Chernoff's Bound

If $\tilde{x}$ is a random variable with moment generating function $M_{\tilde{x}}(t) = E[e^{t\tilde{x}}]$, then, for $a > 0$, we have

$$P(\tilde{x} \geq a) \leq \inf_{t \geq 0} e^{-ta} M_{\tilde{x}}(t) \leq e^{-ta} M_{\tilde{x}}(t) \quad \forall t > 0$$

$$(P(\tilde{x} \leq a) \leq e^{-ta} M_{\tilde{x}}(t) \quad \forall t < 0) \qquad \rightarrow \text{exercise}$$

$\underline{Proof:}$

$$\begin{aligned} t > 0: \quad P(\tilde{x} \geq a) &= P(e^{t\tilde{x}} \geq e^{ta}) \quad (\because t > 0) \\ &\leq \frac{E[e^{t\tilde{x}}]}{e^{ta}} = e^{-ta} M_{\tilde{x}}(t) \end{aligned}$$

<u><Homework></u>. Derive the tightest Chernoff's Bound for Poisson random variable $\tilde{x} \sim P(x; \lambda)$.

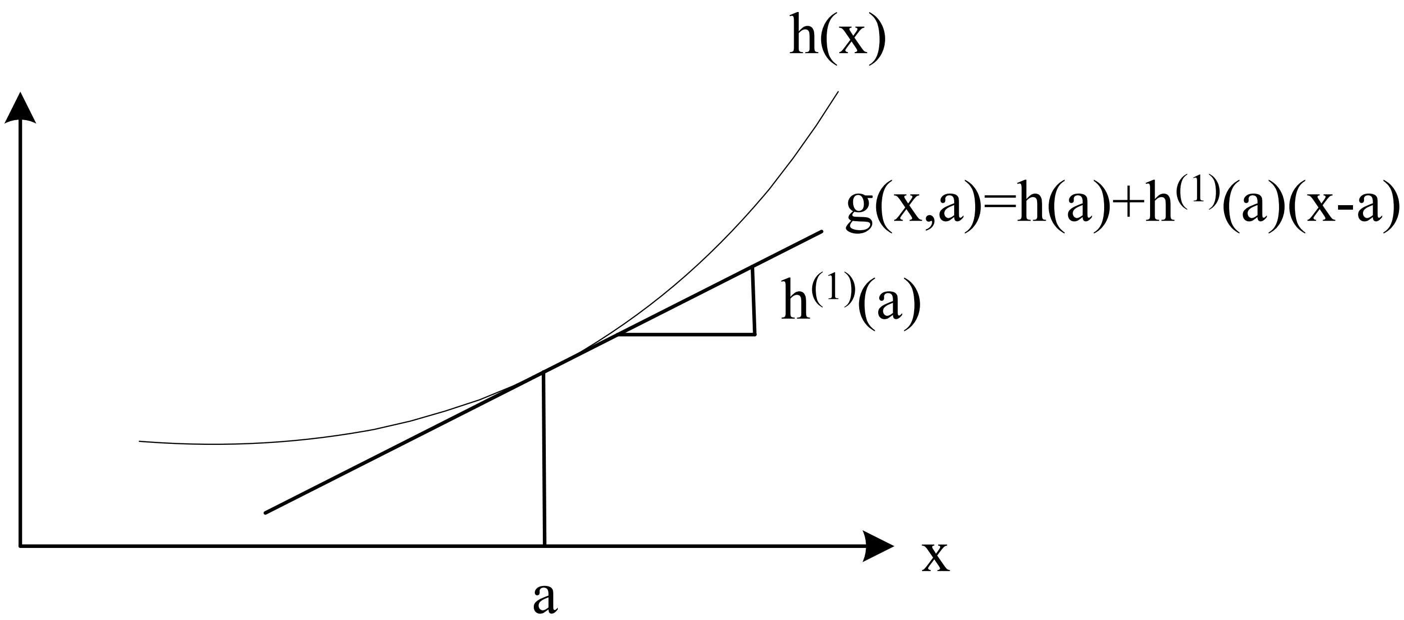
Jensen's Inequality

Lemma. Let h be a convex function. Define the linear function g that is tangent to h at the point a as follows:

$$g(x, a) \stackrel{\text{def}}{=} h(a) + h^{(1)}(a)(x - a).$$

Then,

$$g(x, a) \leq h(x), \quad \text{for all } x.$$



Jensen's Inequality

Jensen's Inequality. If h is a differentiable convex function, defined on real variables, then

$$E[h(\tilde{x})] \geq h(E[\tilde{x}]).$$

Proof:

From the previous lemma, we have

$$h(\tilde{x}) \geq h(a) + h'(a)(\tilde{x} - a)$$

Let $a = E[\tilde{x}]$. Taking $E[\,]$ on both sides yields

$$\begin{aligned} E[h(\tilde{x})] &\geq h(E[\tilde{x}]) + h'(a)[E[\tilde{x}] - E[\tilde{x}]] \\ &= h(E[\tilde{x}]) \end{aligned}$$

Limit Theorems

Theorem (Weak Law of Large Numbers): Let $\tilde{S}_n = \tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_n$, where $\tilde{x}_1, \tilde{x}_2, \ldots \tilde{x}_n, \ldots$ are i.i.d. random variables with $\underline{finite}$ mean $E[\tilde{x}]$, then for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|\frac{\tilde{S}_n}{n} - E[\tilde{x}]| \geq \varepsilon) = 0$$

Theorem (Strong Law of Large Numbers): Let $\tilde{S}_n = \tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_n$, where $\tilde{x}_1, \tilde{x}_2, \ldots \tilde{x}_n, \ldots$ are i.i.d. random variables with $\underline{finite}$ mean $E[\tilde{x}]$, then for any $\varepsilon > 0$,

$$P(\lim_{n\to\infty} |\frac{\tilde{S}_n}{n} - E[\tilde{x}]| \geq \varepsilon) = 0$$

Limit Theorems

Theorem (Central Limit Theorem): Let $\tilde{S}_n = \tilde{x}_1 + \tilde{x}_2 + \ldots + \tilde{x}_n$, where $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n$ are i.i.d. random variables with finite mean $E[\tilde{x}]$ and finite variance $\sigma^2_{\tilde{x}} < \infty$, then,

$$\begin{aligned}
\lim_{n\to\infty} P\left(\frac{\tilde{S}_n - nE[\tilde{x}]}{\sqrt{n}\sigma} \le y\right) &= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
&\sim N(0,1)
\end{aligned}$$

Normalized Gaussian distribution

Strong Law of Large Numbers

To motivate our discussion we perform a simple coin tossing experiment. Consider a coin that lands heads up with probability p. For the sake of the example let $p = 1/4$ and assign the value of 1 to heads and 0 to tails. An experiment consists of an infinite number of tosses. Let Ω denote the set of outcomes of all possible experiments. For any particular experiment $\omega \in \Omega$ the corresponding sequence of 0s and 1s is known deterministically and is termed the *sample path* of ω. The "randomness" in experiments arises by selecting one experiment from the set. We define $Y_n(\omega)$ to be the statistical average of the first n outcomes of experiment ω. Intuitively, for large n the value of $Y_n(\omega)$ is close to p since heads lands up with probability p. Clearly, however, there are sample paths for which this is not the case, and we find it convenient to list two such paths.

Strong Law of Large Numbers

The sample path of experiment ω_1 consists of an alternation of heads and tails (starting by 1,0,1,0, ...). An easy calculation shows that

$$Y_n(\omega_1) = \begin{cases} 1/2, & n = 2k, \\ k/(2k-1), & n = 2k-1, \quad k = 1, 2, \ldots \end{cases} .$$

Notice that $Y_n(\omega_1)$ converges to $1/2$ and not to p as expected.

Strong Law of Large Numbers

Definition 5.39 is an adaptation of the definition of convergence for a deterministic sequence that accounts for the fact that the sequence arises from a random experiment. To see this, recall that by definition of a limit, if a deterministic sequence a_n satisfies $a_n \to a$ then, for any $\epsilon > 0$, there exists a value n(ϵ) so that $|a_n - a| \leq \epsilon$ as $n \to \infty$ for all $n \geq n(\epsilon)$. There are thus *no occurrences* of $|a_n - a| > \epsilon$ for values of $n \geq n(\epsilon)$. When sequences correspond to random experiments, as in the coin tossing experiment mentioned earlier, this type of convergence is too strong. For example, the sample path of ω_1 converges to a value different from p and the sample path of ω_2 does not converge. There is still a sense of convergence to p, however, since the set of experiments that converge to p have probability 1. Violations of sample paths that do not converge to p have probability of 0.

Strong Law of Large Numbers

Recall that since a probability of 0 does not imply impossibility (see Section 4.5.2) we can only conclude that violations of this type of convergence are extremely rare but not impossible.

We can now state the strong law of large numbers as

$$P\left[\ \lim_{n\to\infty}\left|\frac{S_n}{n}-E[X]\right|\geq\epsilon\ \right]=0$$

which can equivalently be stated, using Definition 5.39, as

$$\frac{S_n}{n}\rightarrow E[X]\ \ as\ \ n\rightarrow\infty,\ \ \textbf{Strong Law of Large Numbers}.$$

Strong Law of Large Numbers

The strong law makes a precise statement regarding sample paths obtained in the "typical" experiment, that is, for sufficiently large n there is a large probability that a randomly selected sample path has a statistical average close to $E[X]$. In contrast, the weak law makes a statement regarding the entire ensemble of sample paths, that is, for sufficiently large n there is a large probability that, *averaged over all sample paths*, the statistical average is close to $E[X]$. The weak law does not make any statement regarding particular sample paths of random experiments and, specifically, does not imply that a randomly selected sample path converges to $E[X]$. Conceivably it could be the case that all sample paths either converge to values different from $E[X]$ or do not converge at all (as with experiments ω_1 and ω_2, respectively) and the weak law could still hold. In these cases the strong law would be violated. It is obvious from what we have said that the strong law implies the weak law.