# 6. Stochastic Processes and Limiting Distributions

# 6.1. Stochastic processes

- **Definition.** A stochastic process is a sequence (a vector) of random variables $\{\tilde{x}_t\}_{t=0}^{T}$ defined on the probability space $(\Omega, \mathcal{F}, P)$.
- *Note:* More generally, we can define a stochastic process of random objects $\{\tilde{x}_t\}_{t=0}^{T}$, where $\tilde{x}_t : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega', \mathcal{F}')$.
- *Note:* We can have $T = \infty$ in all the previous and forthcoming expressions.
- **Definition.** A stochastic process $\{\tilde{x}_t\}_{t=0}^{\infty}$ defined on the probability space $(\Omega, \mathcal{F}, P)$ is strongly (or strictly) stationary if the distributions

$$P_{\tilde{x}_t, \tilde{x}_{t+1}, \tilde{x}_{t+2}, \ldots, \tilde{x}_{t+n}}(C) = P\left\{(\tilde{x}_t, \tilde{x}_{t+1}, \tilde{x}_{t+2}, \ldots, \tilde{x}_{t+n}) \in C\right\}$$

do not depend on $t$, for all non-negative integers $n = 0, 1, 2, \ldots$, and all $C \in \mathcal{B}\left(\mathbb{R}^{n+1}\right)$.

- **Definition.** A stochastic process $\{\tilde{x}_t\}_{t=0}^{\infty}$ on the probability space $(\Omega, \mathcal{F}, P)$ with $\tilde{x}_t \in \mathrm{L}^2$ for all $t$ is weakly stationary if

$$\mathrm{E}\left(\tilde{x}_t\right) = \mu \text{ for } t = 0, 1, \ldots$$

and $\mathrm{Cov}\left(\tilde{x}_t, \tilde{x}_{t+n}\right)$ depends only on $n$, for all $t$, i.e.,

$$\mathrm{Cov}\left(\tilde{x}_t, \tilde{x}_{t+n}\right) = \hat{\gamma}(n), \quad \text{for } t = 0, 1, \ldots \text{ and } n = 0, 1, \ldots$$

- The covariance $\mathrm{Cov}\left(\tilde{x}_t, \tilde{x}_{t+n}\right) \equiv \gamma(t, n)$ is called autocovariance since it is the covariance between the members of a single stochastic process.
- Note that $\gamma(t, n) = \hat{\gamma}(n)$ for all $t$ when the stochastic process is weakly stationary. Moreover, in this case $\mathrm{Var}\left(\tilde{x}_t\right) = \hat{\gamma}(0)$ for all $t$.

- **Definition.** The random variables $\{\tilde{x}_t\}_{t=0}^{\infty}$ on the probability space $(\Omega, \mathcal{F}, P)$ are independent if the random variables of any finite subset of $\{\tilde{x}_t\}_{t=0}^{\infty}$ are independent.
- **Definition.** The random variables $\{\tilde{x}_t\}_{t=0}^{T}$ on the probability space $(\Omega, \mathcal{F}, P)$ are identically distributed if they have the same (marginal) distribution.
- We say that the random variables $\{\tilde{x}_t\}_{t=0}^{T}$ on the probability space $(\Omega, \mathcal{F}, P)$ are i.i.d. (identically and independently distributed) or that the stochastic process is i.i.d. when they simultaneously satisfy the previous two definitions.
- *Note:* i.i.d. processes are strongly stationary.

- **Definition.** The stochastic process $\{\tilde{x}_t\}_{t=0}^T$ on the probability space $(\Omega, \mathcal{F}, P)$ with $\tilde{x}_t \in \mathrm{L}^1$ for all $t$ is white noise if it is i.i.d. and $\mathrm{E}(\tilde{x}_t) = 0$ for all $t$.

- **Definition.** The stochastic process $\{\tilde{x}_t\}_{t=0}^T$ on the probability space $(\Omega, \mathcal{F}, P)$ is a random walk if $\tilde{x}_{t+1} = \tilde{x}_t + \tilde{y}_{t+1}$ for $t = 0, 1, \ldots$, where $\{\tilde{y}_t\}_{t=1}^T$ is i.i.d. and the value $x_0$ taken by the random variable $\tilde{x}_0$ is exogenously given.

- If the stochastic process $\{\tilde{x}_t\}_{t=0}^T$ on the probability space $(\Omega, \mathcal{F}, P)$ is a random walk with $\tilde{x}_{t+1} = \tilde{x}_t + \tilde{y}_{t+1}$ and $\mathrm{E}(\tilde{y}_t) = \mu \neq 0$ for all $t$, then $\{\tilde{x}_t\}_{t=0}^T$ is called a random walk with drift $\mu$. If $\mathrm{E}(\tilde{y}_t) = 0$ for all $t$, then $\{\tilde{x}_t\}_{t=0}^T$ is called a (symmetric) random walk.

- Thus, a random walk $\{\tilde{x}_t\}_{t=0}^T$ with drift $\mu$ can be written as $\tilde{x}_{t+1} = \mu + \tilde{x}_t + \tilde{\varepsilon}_{t+1}$ , where $\{\tilde{\varepsilon}_t\}_{t=1}^T$ is white noise since $\tilde{\varepsilon}_t \equiv \tilde{y}_t - \mu$ for all $t$.

# 6.2. Filtrations and martingales

- **Definition.** A filtration $\mathbb{F}$ on $(\Omega, \mathcal{F})$ is a sequence of $\sigma$-algebras, $\mathbb{F} = \{\mathcal{F}_t\}_{t=0}^{T}$ such that $\mathcal{F}_t \subset \mathcal{F}_{t+1} \subset \mathcal{F}$ for all $t$.
- **Definition.** If $\mathbb{F} = \{\mathcal{F}_t\}_{t=0}^{T}$ is a filtration on $(\Omega, \mathcal{F})$ and $\{\tilde{x}_t\}_{t=0}^{T}$ is a stochastic process on $(\Omega, \mathcal{F}, P)$ such that $\tilde{x}_t$ is $\mathcal{F}_t$-measurable for $t = 0, 1, \dots T$, then we say that the stochastic process $\{\tilde{x}_t\}_{t=0}^{T}$ is adapted to $\mathbb{F}$.
- The **natural filtration** associated to the stochastic process $\{\tilde{x}_t\}_{t=0}^{T}$ on $(\Omega, \mathcal{F}, P)$ is the simplest filtration which the stochastic process $\{\tilde{x}_t\}_{t=0}^{T}$ is adapted to. All information concerning the process, and only that information, is available in the natural filtration.

- Formally, the natural filtration $\mathbb{F}^X = \{\mathcal{F}_t^X\}_{t=0}^T$ associated to the stochastic process $X = \{\tilde{x}_t\}_{t=0}^T$ on $(\Omega, \mathcal{F}, P)$ is given by

$$\mathcal{F}_t^X = \sigma\left\{ \tilde{x}_j^{-1}(B) \middle| \; j \leq t, \; B \in \mathcal{B} \right\},$$

  i.e., $\mathcal{F}_t^X$ is the smallest $\sigma$-algebra on $\Omega$ that contains all pre-images of the Borel sets of $\mathbb{R}$ for all $j$ up to $t$.

- Any stochastic process is adapted to its natural filtration.

- **Definition.** Let $\{\tilde{x}_t\}_{t=0}^T$ be a stochastic process on $(\Omega, \mathcal{F}, P)$ adapted to the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t=0}^T$ on $(\Omega, \mathcal{F})$ and $\tilde{x}_t \in \mathrm{L}^1$ for all $t$. Then, we say that the stochastic process $\{\tilde{x}_t\}_{t=0}^T$ is a martingale (submartingale) (supermartingale) relative to $\mathbb{F}$ if,
$\mathrm{E}(\tilde{x}_{t+1} | \mathcal{F}_t) = (\geq)(\leq)\, \tilde{x}_t$ a.s. $[P]$, for $t = 0, 1, \ldots T-1$.

- When we say that $X = \{\tilde{x}_t\}_{t=0}^T$ is a martingale (submartingale) (supermartingale) without mentioning the filtration, we mean that $X = \{\tilde{x}_t\}_{t=0}^T$ is a martingale (submartingale) (supermartingale) relative to its associated natural filtration $\mathbb{F}^X$.

- **Proposition.** The stochastic process $X = \{\tilde{x}_t\}_{t=0}^{T}$ on $(\Omega, \mathcal{F}, P)$ is a martingale (submartingale) (supermartingale) relative to its associated natural filtration $\mathbb{F}^X = \left\{\mathcal{F}_t^X\right\}_{t=0}^{T}$ if and only if

$$\mathrm{E}\left(\tilde{x}_{t+1} \left| \tilde{x}_t, \tilde{x}_{t-1}, ..., \tilde{x}_1, \tilde{x}_0 \right.\right) = (\geq)(\leq)\tilde{x}_t \ ,$$

$$\text{a.s. } [P], \text{ for } t = 0, 1, ..., T-1, \qquad (*)$$

- **Proof.** Just note that

$$\mathrm{E}\left(\tilde{x}_{t+1} \left| \tilde{x}_t, \tilde{x}_{t-1}, ..., \tilde{x}_0 \right.\right) = \mathrm{E}\left(\tilde{x}_{t+1} \left| \mathcal{F}_t^X \right.\right), \text{ a.s. } [P]. \qquad Q.E.D.$$

- Note that the condition $(*)$ in the previous proposition is equivalent to the following:

$$\mathrm{E}\left(\tilde{x}_{t+1} \left| \tilde{x}_t = x_t, \tilde{x}_{t-1} = x_{t-1}, ..., \tilde{x}_0 = x_0 \right.\right) = (\geq)(\leq)\, x_t,$$

$$\text{a.s. } \left[P_{\tilde{x}_t, \tilde{x}_{t-1}, ..., \tilde{x}_0}\right], \text{ for } t = 0, 1, ..., T-1.$$

- A random walk with no drift (with positive drift) (with negative drift) is a martingale (submartingale) (supermartingale) relative to its associated natural filtration. Recall that every stochastic process is adapted to its natural filtration.

- **Proposition.** If the stochastic process $\{\tilde{x}_t\}_{t=0}^{T}$ on $(\Omega, \mathcal{F}, P)$ adapted to the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t=0}^{T}$ on $(\Omega, \mathcal{F})$ is a martingale (submartingale) (supermartingale) relative to $\mathbb{F} = \{\mathcal{F}_t\}_{t=0}^{T}$ then

  **(a)**

  $$\mathrm{E}\left(\tilde{x}_{t+1}\left|\tilde{x}_t, \tilde{x}_{t-1}, ..., \tilde{x}_1, \tilde{x}_0\right.\right) = (\geq)(\leq)\tilde{x}_t \text{ , a.s. } [P], \text{ for } t = 0, 1, ..., T-1,$$

  i.e., $\{\tilde{x}_t\}_{t=0}^{T}$ is a martingale relative to its natural filtration $\mathbb{F}^X$; and

  **(b)**

  $$\mathrm{E}\left(\tilde{x}_{t+1}\left|\tilde{x}_t = x_t, \tilde{x}_{t-1} = x_{t-1}, ..., \tilde{x}_0 = x_0\right.\right) = (\geq)\,(\leq)\,x_t,$$

  $$\text{a.s. } \left[P_{\tilde{x}_t, \tilde{x}_{t-1}, ..., \tilde{x}_0}\right], \text{ for } t = 0, 1, ..., T-1.$$

- **Proof. (a)** Recall that $\mathcal{F}(\tilde{x}_t)$ is the $\sigma-$algebra on $\Omega$ induced by the random variable $\tilde{x}_t$, i.e., the coarsest $\sigma-$algebra that makes $\tilde{x}_t$ measurable. Observe that $\mathcal{F}(\tilde{x}_t) \subset \mathcal{F}_t^X \subset \mathcal{F}_t$ so that $\tilde{x}_t$ is $\mathcal{F}_t^X-$measurable. Therefore,

$$\mathrm{E}\left(\tilde{x}_{t+1}\left|\tilde{x}_t, \tilde{x}_{t-1}, ..., \tilde{x}_0\right.\right) = \mathrm{E}\left(\tilde{x}_{t+1}\left|\mathcal{F}_t^X\right.\right) = \mathrm{E}\left(\mathrm{E}\left(\tilde{x}_{t+1}\left|\mathcal{F}_t\right.\right)\left|\mathcal{F}_t^X\right.\right)$$

$$= (\geq)(\leq)\mathrm{E}\left(\tilde{x}_t\left|\mathcal{F}_t^X\right.\right) = \tilde{x}_t, \text{ a.s. } [P],$$

where the first equality follows from the definition of $\mathcal{F}_t^X$, the second equality follows from the law of iterated expectations, the third equality (inequality) follows from the martingale (submartingale) (supermartingale) assumption, $\mathrm{E}\left(\tilde{x}_{t+1}\left|\mathcal{F}_t\right.\right) = (\geq)(\leq)\,\tilde{x}_t$, and the fourth equality follows from the $\mathcal{F}_t^X-$measurability of $\tilde{x}_t$.
**(b)** Obvious since (a) and (b) are equivalent. *Q.E.D.*

- **Corollary 1.** If the stochastic process $\{\tilde{x}_t\}_{t=0}^{T}$ on $(\Omega, \mathcal{F}, P)$ is a martingale (submartingale) (supermartingale) relative to the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t=0}^{T}$, then

$$\mathrm{E}\left(\tilde{x}_{t+1}\left|\tilde{x}_t\right.\right) = (\geq)\,(\leq)\,\tilde{x}_t, \text{ a.s. } [P], \text{ for } t = 0, 1, ..., T-1.$$

and

$$\mathrm{E}\left(\tilde{x}_{t+1}\left|\tilde{x}_t = x_t\right.\right) = (\geq)\,(\leq)\,x_t, \text{ a.s. } \left[P_{\tilde{x}_t}\right], \text{ for } t = 0, 1, ..., T-1.$$

- **Proof.** Note that $\mathcal{F}(\tilde{x}_t) \subset \mathcal{F}_t$, where $\mathcal{F}(\tilde{x}_t)$ is the $\sigma-$algebra on $\Omega$ induced by the random variable $\tilde{x}_t$, and, hence,

$$\mathrm{E}\left(\tilde{x}_{t+1}\left|\tilde{x}_t\right.\right) = \mathrm{E}\left(\tilde{x}_{t+1}\left|\mathcal{F}(\tilde{x}_t)\right.\right) = \mathrm{E}\left(\mathrm{E}\left(\tilde{x}_{t+1}\left|\mathcal{F}_t\right.\right)\left|\mathcal{F}(\tilde{x}_t)\right.\right)$$
$$= (\geq)\,(\leq)\,\mathrm{E}\left(\tilde{x}_t\left|\mathcal{F}(\tilde{x}_t)\right.\right) = \tilde{x}_t, \text{ a.s. } [P],$$

where the second equality follows from the law of iterated expectations and the third equality (inequality) follows from the martingale (submartingale) (supermartingale) assumption, $\mathrm{E}\left(\tilde{x}_{t+1}\left|\mathcal{F}_t\right.\right) = (\geq)\,(\leq)\,\tilde{x}_t$. *Q.E.D.*

- **Corollary 2.** If the stochastic process $\{\tilde{x}_t\}_{t=0}^T$ on $(\Omega, \mathcal{F}, P)$ is a martingale relative to the filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t=0}^T$, then

$$\mathrm{E}\left(\tilde{x}_{t+n}\left|\mathcal{F}_t\right.\right) = \tilde{x}_t, \quad \text{a.s. } [P], \text{ for } t = 0, 1, \ldots T-1 \text{ and } n = 1, 2, \ldots$$

and

$$\mathrm{E}\left(\tilde{x}_{t+n}\left|\tilde{x}_t = x_t\right.\right) = x_t, \quad \text{a.s. } \left[P_{\tilde{x}_t}\right], \text{ for } t = 0, 1, \ldots T-1 \text{ and } n = 1, 2, \ldots$$

- **Proof.** Since $\mathcal{F}_t \subset \mathcal{F}_{t+1} \subset \ldots \subset \mathcal{F}_{t+n-1}$, we apply the law of iterated expectations and the martingale property,

$$\mathrm{E}\left(\tilde{x}_{t+n}\left|\mathcal{F}_t\right.\right) = \mathrm{E}\left(\mathrm{E}\left(\tilde{x}_{t+n}\left|\mathcal{F}_{t+n-1}\right.\right)\left|\mathcal{F}_t\right.\right) = \mathrm{E}\left(\tilde{x}_{t+n-1}\left|\mathcal{F}_t\right.\right),$$

and we get the desired result by induction, i.e., by repeating the previous procedure $n$ times until we get

$$\mathrm{E}\left(\tilde{x}_{t+n}\left|\mathcal{F}_t\right.\right) = \mathrm{E}\left(\tilde{x}_t\left|\mathcal{F}_t\right.\right) = \tilde{x}_t. \qquad \textit{Q.E.D.}$$

- *Notation:* In Dynamic Economics we use the notation $\mathrm{E}_t\left(\tilde{x}_j\right)$ to mean $\mathrm{E}\left(\tilde{x}_j \left| \mathcal{F}_t \right.\right)$.

- If the stochastic process $\{\tilde{x}_t\}_{t=0}^{\infty}$ is "independently distributed" then $\mathrm{E}_t\left(\tilde{x}_{t+j}\right) = \mathrm{E}\left(\tilde{x}_{t+j}\right)$ for all $t$ and $j > 0$.

- If the stochastic process $\{\tilde{x}_t\}_{t=0}^{\infty}$ is "identically distributed" then $\mathrm{E}\left(\tilde{x}_t\right) = \mathrm{E}\left(\tilde{x}\right)$ for all $t$, where $\tilde{x}$ is a random variable having the same distribution as the random variable $\tilde{x}_t$ for all $t$.

- Thus, if the stochastic process $\{\tilde{x}_t\}_{t=0}^{\infty}$ is "identically and independently distributed (i.i.d.)" then $\mathrm{E}_t\left(\tilde{x}_{t+j}\right) = \mathrm{E}\left(\tilde{x}\right)$ for all $t$ and $j > 0$, where $\tilde{x}$ is a random variable having the same distribution as the random variable $\tilde{x}_s$ for all $s$.

- Obviously, $\mathrm{E}_t\left(\tilde{x}_{t-j}\right) = \tilde{x}_{t-j}$, for all $t$ and $j \geq 0$.

# 6.3. Convergence in probability, in mean square, in distribution, and almost sure convergence

- Let $\{\tilde{x}_n\}_{n=1}^{\infty}$ be a stochastic process or sequence of random variables on $(\Omega, \mathcal{F}, P)$ . Note that we are making $n = t + 1$.

• **Convergence in probability.** The sequence of random variables $\{\tilde{x}_n\}_{n=1}^{\infty}$ is said to converge to the random variable $\tilde{x}$ in probability if

$$\lim_{n\to\infty} P\{|\tilde{x}_n - \tilde{x}| \geq \varepsilon\} \equiv \lim_{n\to\infty} P\{\omega \in \Omega \,|\, |\tilde{x}_n(\omega) - \tilde{x}(\omega)| \geq \varepsilon\} = 0, \text{ for all } \varepsilon > 0,$$

or

$$\lim_{n\to\infty} P\{|\tilde{x}_n - \tilde{x}| < \varepsilon\} \equiv \lim_{n\to\infty} P\{\omega \in \Omega \,|\, |\tilde{x}_n(\omega) - \tilde{x}(\omega)| < \varepsilon\} = 1, \text{ for all } \varepsilon > 0.$$

- This concept is the same as that of "convergence in measure".
- We use the notation $\tilde{x}_n \xrightarrow{p} \tilde{x}$ or $\underset{n\to\infty}{\text{plim}}\ \tilde{x}_n = \tilde{x}$.

- **Convergence in mean square.** The sequence of random variables $\{\tilde{x}_n\}_{n=1}^{\infty}$ belonging to $\mathrm{L}^2$ is said to converge to the random variable $\tilde{x} \in \mathrm{L}^2$ in mean square if

$$\lim_{n\to\infty} \mathrm{E}\left[(\tilde{x}_n - \tilde{x})^2\right] = 0.$$

- We write $\tilde{x}_n \xrightarrow{m} \tilde{x}$.

- **Proposition.** If $\lim_{n\to\infty} \mathrm{E}(\tilde{x}_n) = c$, where $c$ is a constant, and $\lim_{n\to\infty} \mathrm{Var}(\tilde{x}_n) = 0$, then $\tilde{x}_n \xrightarrow{m} c$.

- **Proof.**

$$\begin{aligned} \lim_{n\to\infty} \mathrm{Var}(\tilde{x}_n) &= \lim_{n\to\infty} \mathrm{Var}(\tilde{x}_n - c) \\ &= \lim_{n\to\infty} \mathrm{E}\left[(\tilde{x}_n - c)^2\right] - \lim_{n\to\infty} \left[\mathrm{E}(\tilde{x}_n - c)\right]^2 \\ &= \lim_{n\to\infty} \mathrm{E}\left[(\tilde{x}_n - c)^2\right] - \underbrace{\left[\lim_{n\to\infty} \mathrm{E}(\tilde{x}_n) - c\right]^2}_{=0} \\ &= \lim_{n\to\infty} \mathrm{E}\left[(\tilde{x}_n - c)^2\right] = 0. \qquad \textit{Q.E.D.} \end{aligned}$$

- **Corollary.** If $\mathrm{E}(\tilde{x}_n) = c$ for all $n$, where $c$ is a constant, and $\lim_{n\to\infty} \mathrm{Var}(\tilde{x}_n) = 0$, then $\tilde{x}_n \xrightarrow{m} c$.

- **Almost sure convergence.** The sequence of random variables $\{\tilde{x}_n\}_{n=1}^{\infty}$ is said to converge to the random variable $\tilde{x}$ almost surely (or strongly) if

$$P\left\{\lim_{n\to\infty}\tilde{x}_n=\tilde{x}\right\}\equiv P\left\{\omega\in\Omega\,\middle|\,\lim_{n\to\infty}\tilde{x}_n(\omega)=\tilde{x}(\omega)\right\}=1,$$

  or

$$P\left\{\lim_{n\to\infty}\tilde{x}_n\neq\tilde{x}\right\}\equiv P\left\{\omega\in\Omega\,\middle|\,\lim_{n\to\infty}\tilde{x}_n(\omega)\neq\tilde{x}(\omega)\right\}=0.$$

- This concept is the same as that of "convergence almost everywhere".
- We write $\tilde{x}_n \xrightarrow{a.s} \tilde{x}$.

- **Sure convergence.** The sequence of random variables $\{\tilde{x}_n\}_{n=1}^{\infty}$ is said to converge to the random variable $\tilde{x}$ surely if

$$\lim_{n\to\infty} \tilde{x}_n = \tilde{x}, \text{ that is, } \lim_{n\to\infty} \tilde{x}_n(\omega) = \tilde{x}(\omega) \text{ for all } \omega \in \Omega.$$

- This concept is the same as that of "pointwise convergence" and is rarely used in Statistics.
- We write $\tilde{x}_n \xrightarrow{s} \tilde{x}$.

- **Convergence in distribution.** The sequence of random variables $\{\tilde{x}_n\}_{n=1}^{\infty}$ is said to converge to the random variable $\tilde{x}$ in distribution (or weakly) if the distribution function $F_n$ of $\tilde{x}_n$ converges pointwise to the distribution $F$ of $\tilde{x}$ at every continuity point of $F$.

- We write $\tilde{x}_n \xrightarrow{d} \tilde{x}$, $\tilde{x}_n \longrightarrow F$, $\tilde{x}_n \longrightarrow \widehat{P}$, $\tilde{x}_n \overset{a}{\sim} \tilde{x}$ ($\tilde{x}_n$ is asymptotically distributed as $\tilde{x}$), $\tilde{x}_n \overset{a}{\sim} F$ or $\tilde{x}_n \overset{a}{\sim} \widehat{P}$, where $F$ and $\widehat{P}$ are called the asymptotic (or limiting) distribution function and the asymptotic distribution of $\{\tilde{x}_n\}_{n=1}^{\infty}$, respectively. Obviously, $\widehat{P}$ is the distribution associated with the distribution function $F$.

- All the previous convergence concepts can be applied to sequences of random vectors, $\{\tilde{X}_n\}_{n=1}^{\infty}$ with $\tilde{X}_n : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. For convergence in probability we just need to replace the absolute value $|\tilde{x}_n(\omega) - \tilde{x}(\omega)|$ by the Euclidean distance $\|\tilde{X}_n(\omega) - \tilde{X}(\omega)\|$, where $\tilde{X}$ is the limiting random vector. For convergence in mean square, we just need to replace $\lim_{n\to\infty} \mathrm{E}\left[(\tilde{x}_n - \tilde{x})^2\right] = 0$ by $\lim_{n\to\infty} \mathrm{E}\left[\|\tilde{X}_n - \tilde{X}\|^2\right] = 0$.

$$\begin{array}{ccccc} & & s & & \\ & & \Downarrow & & \\ & & a.s & & \\ & & \Downarrow & & \\ m & \Longrightarrow & p & \Longrightarrow & d \\ & & & \underset{\text{if } \tilde{x} \text{ is a constant}}{\Longleftarrow} & \end{array}$$

- **Proof of** $s \Longrightarrow a.s$ **(trivial).**
- **Proof of** $a.s \Longrightarrow p$ **(trivial):**

$$P\left\{\lim_{n\to\infty} \tilde{x}_n = \tilde{x}\right\} = 1 \Longrightarrow \lim_{n\to\infty} P\left\{|\tilde{x}_n - \tilde{x}| < \varepsilon\right\} = 1.$$

- **Proof of** $p \Longrightarrow d$ (see the handout).
- **Proof of** $d \Longrightarrow p$ **when** $\tilde{x}$ **is a constant** (see the handout).

- **Proof of $m \Longrightarrow p$ (easy):**
- Recall the Markov inequality: If $\tilde{y}$ is a non-negative random variable, then

$$P\left\{\tilde{y} \geq a\right\} \leq \frac{\mathrm{E}\left(\tilde{y}\right)}{a} \text{ for all } a > 0.$$

Make $a = \varepsilon^2 > 0$ and $\tilde{y} = \left(\tilde{x}_n - \tilde{x}\right)^2 \geq 0$ so that

$$P\left\{\left(\tilde{x}_n - \tilde{x}\right)^2 \geq \varepsilon^2\right\} \leq \frac{\mathrm{E}\left[\left(\tilde{x}_n - \tilde{x}\right)^2\right]}{\varepsilon^2} \text{ for all } \varepsilon \neq 0.$$

Therefore,

$$\lim_{n\to\infty} P\left\{\left(\tilde{x}_n - \tilde{x}\right)^2 \geq \varepsilon^2\right\} \leq \lim_{n\to\infty} \frac{\mathrm{E}\left[\left(\tilde{x}_n - \tilde{x}\right)^2\right]}{\varepsilon^2} \text{ for all } \varepsilon \neq 0.$$

If $\tilde{x}_n \xrightarrow{m} \tilde{x}$, then

$$\lim_{n\to\infty} \mathrm{E}\left[(\tilde{x}_n - \tilde{x})^2\right] = 0$$

so that

$$0 \leq \lim_{n\to\infty} P\left\{(\tilde{x}_n - \tilde{x})^2 \geq \varepsilon^2\right\} \leq \lim_{n\to\infty} \frac{\mathrm{E}\left[(\tilde{x}_n - \tilde{x})^2\right]}{\varepsilon^2} = 0 \text{ for all } \varepsilon \neq 0.$$

Thus,

$$\lim_{n\to\infty} P\left\{|\tilde{x}_n - \tilde{x}| \geq \varepsilon\right\} = 0 \text{ for all } \varepsilon > 0,$$

and, hence, $\tilde{x}_n \xrightarrow{p} \tilde{x}$. *Q.E.D.*

# 6.4. Weak convergence of distribution functions and of probability measures

- **Definition.** The sequence of distribution functions $\{F_n\}_{n=1}^{\infty}$ on $\mathbb{R}^k$ converges weakly to the distribution function $F$ if $F_n$ converges pointwise to $F$ at every continuity point of $F$, i.e., $\lim_{n\to\infty} F_n(x) = F(x)$ if $F$ is continuous at $x \in \mathbb{R}^k$.

- We write $F_n \xrightarrow{w} F$ for weak convergence, while $F_n \longrightarrow F$ denotes pointwise convergence of $F$, i.e., $\lim_{n\to\infty} F_n(x) = F(x)$ for all $x \in \mathbb{R}^k$.

- Therefore, if $\tilde{x}_n \sim F_n(x)$ and $\tilde{x} \sim F(x)$, then $\tilde{x}_n \xrightarrow{d} \tilde{x} \iff F_n \xrightarrow{w} F$.

- **Example:**

$$\tilde{x}_n \sim F_n(x) = \begin{cases} 0 & \text{if } x < \dfrac{1}{n} \\ 1 & \text{if } x \geq \dfrac{1}{n} \end{cases}$$

and

$$\tilde{x} \sim F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0. \end{cases}$$

- Then,

$$\lim_{n\to\infty} F_n(x) = F(x) \text{ for all } x \neq 0,$$

and

$$\lim_{n\to\infty} F_n(0) = 0 \neq F(0) = 1.$$

Therefore,

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0. \end{cases}$$

However, $F$ is discontinuous at 0. Hence, $F_n \xrightarrow{w} F \iff \tilde{x}_n \xrightarrow{d} \tilde{x}$.

- **Definition.** The sequence of measures $\{\mu_n\}_{n=1}^{\infty}$ on $(\mathbb{R}^k, \mathcal{B})$ converges weakly to the measure $\mu$ if $\lim_{n\to\infty} \mu_n(B) = \mu(B)$ for all Borel sets $B \in \mathcal{B}(\mathbb{R}^k)$ such that $\mu(\partial B) = 0$, where $\partial B$ is the boundary of $B$.

- We write $\mu_n \xrightarrow{w} \mu$.

- **Proposition.** $F_n \xrightarrow{w} F$ if and only if $P_n \xrightarrow{w} \widehat{P}$, where $P_n$ and $\widehat{P}$ are the distributions (probability measures on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$) associated with $F_n$ and $F$, respectively.

- *Note:* To say that $\lim_{n\to\infty} F_n(x) = F(x)$ when $F$ is continuous at $x \in \mathbb{R}$ is equivalent to say that

$$\lim_{n\to\infty} P_n(-\infty, x] = \widehat{P}(-\infty, x]$$

  at every continuity point of $F$, i.e., at every point where $\widehat{P}\{x\} = 0$.

- Therefore, $\tilde{x}_n \xrightarrow{d} \tilde{x} \iff F_n \xrightarrow{w} F \iff P_n \xrightarrow{w} \widehat{P}$.

- **Proposition.** The sequence of measures $\{\mu_n\}_{n=1}^{\infty}$ on $\left(\mathbb{R}^k, \mathcal{B}\right)$ converges weakly to the measure $\mu$, $\mu_n \xrightarrow{w} \mu$, if and only if

$$\lim_{n\to\infty} \int_{\mathbb{R}^k} g(x) d\mu_n = \int_{\mathbb{R}^k} g(x) d\mu.$$

  for all bounded and continuous functions $g : \left(\mathbb{R}^k, \mathcal{B}\left(\mathbb{R}^k\right)\right) \longrightarrow \left(\mathbb{R}, \mathcal{B}\right)$.

- **Corollary (Portmanteau theorem).** The sequence of random vectors $\left\{\tilde{X}_n\right\}_{n=1}^{\infty}$, $\tilde{X}_n : \left(\Omega, \mathcal{F}, P\right) \longrightarrow \left(\mathbb{R}^k, \mathcal{B}\left(\mathbb{R}^k\right)\right)$, converges in distribution to the random vector $\tilde{X}$ if and only if

$$\lim_{n\to\infty} \mathrm{E}\left[g\left(\tilde{X}_n\right)\right] = \mathrm{E}\left[g\left(\tilde{X}\right)\right]$$

  for all bounded and continuous functions $g : \left(\mathbb{R}^k, \mathcal{B}\left(\mathbb{R}^k\right)\right) \longrightarrow \left(\mathbb{R}, \mathcal{B}\right)$.

- Recall that $\varphi_{\tilde{y}}(t) = \mathrm{E}\left(e^{it\tilde{y}}\right)$ is the characteristic function, $M_{\tilde{y}}(t) = \mathrm{E}\left(e^{t\tilde{y}}\right)$ is the moment-generating function, and $\Lambda_{\tilde{y}}(t) = \mathrm{E}\left(e^{-t\tilde{y}}\right)$ is the Laplace transform of the random variable $\tilde{y}$ (or of its distribution).

- **Theorem (Lévy's convergence theorem).** $\tilde{x}_n \xrightarrow{d} \tilde{x}$ if and only if $\lim_{n\to\infty} \varphi_{\tilde{x}_n}(t) = \varphi_{\tilde{x}}(t)$ for all $t \in \mathbb{R}$.

  **or**

  $\tilde{x}_n \xrightarrow{d} \tilde{x}$ if and only if $\lim_{n\to\infty} M_{\tilde{x}_n}(t) = M_{\tilde{x}}(t)$ for all $t$ in the domain of $M_{\tilde{x}}$, provided $M_{\tilde{x}}$ is well-defined (i.e., finite) in a neighborhood of $t = 0$.

  **or**

  $\tilde{x}_n \xrightarrow{d} \tilde{x}$ if and only if $\lim_{n\to\infty} \Lambda_{\tilde{x}_n}(t) = \Lambda_{\tilde{x}}(t)$ for all $t$ in the domain of $\Lambda_{\tilde{x}}$, provided $\Lambda_{\tilde{x}}$ is well-defined (i.e., finite) in a neighborhood of $t = 0$.

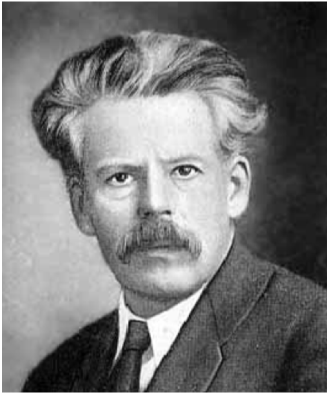
Paul Lévy (1886 - 1971)

- **Continuous mapping theorem.** If a sequence of random vectors $\{\tilde{X}_n\}_{n=1}^{\infty}$, $\tilde{X}_n : (\Omega, \mathcal{F}, P) \longrightarrow \left(\mathbb{R}^k, \mathcal{B}\left(\mathbb{R}^k\right)\right)$, converges in distribution (in probability) (almost surely) to the random vector $\tilde{X}$ and the function $g : \left(\mathbb{R}^k, \mathcal{B}\left(\mathbb{R}^k\right)\right) \longrightarrow \left(\mathbb{R}^m, \mathcal{B}\left(\mathbb{R}^m\right)\right)$ is continuous, then $g(\tilde{X}_n)$ converges in distribution (in probability) (almost surely) to $g\left(\tilde{X}\right)$.

- **Slutsky's theorem.** If the sequence of random variables $\{\tilde{x}_n\}_{n=1}^{\infty}$ converges in distribution to the random variable $\tilde{x}$ and the sequence of random variables $\{\tilde{y}_n\}_{n=1}^{\infty}$ converges in distribution (or in probability) to the constant $c$, then

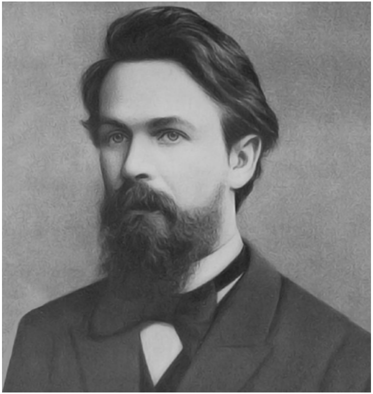
  1) $\tilde{x}_n + \tilde{y}_n \xrightarrow{d} \tilde{x} + c$,
  2) $\tilde{x}_n \cdot \tilde{y}_n \xrightarrow{d} c\tilde{x}$,
  3) $\tilde{x}_n / \tilde{y}_n \xrightarrow{d} \tilde{x}/c$ for $c \neq 0$.

- All the previous discussion about convergence can be extended to a framework where the random variable $\tilde{x}_r$ is characterized by a vector $r \in \mathbb{R}^k$ and $r$ converges to $a \in \overline{\mathbb{R}}^k$.



Evgeny Slutsky (1880 – 1948)

# 6.5. Markov processes



Andrei Markov (1856 - 1922)

- **Discrete Markov Chains.**

- A discrete Markov process (or Markov chain) is a sequence of discrete random objects $\{\tilde{x}_t\}_{t=0}^{\infty}$ defined on the probability space $(\Omega, \mathcal{F}, P)$ taking values on the same countable set $S$ (the state space), $\tilde{x}_t : (\Omega, \mathcal{F}, P) \longrightarrow \left(S, 2^S\right)$, where the conditional distribution of $\tilde{x}_{t+1}$ given $(\tilde{x}_t, \tilde{x}_{t-1}, ..., \tilde{x}_0) = (x_t, x_{t-1}, ..., x_0)$ depends only on the value $x_t$ taken by $\tilde{x}_t$ for $t = 0, 1, ...$, i.e.,

$$P\left\{\tilde{x}_{t+1} = x_{t+1} \middle| \tilde{x}_t = x_t, \tilde{x}_{t-1} = x_{t-1}, ..., \tilde{x}_0 = x_0\right\}$$

$$= P\left\{\tilde{x}_{t+1} = x_{t+1} \middle| \tilde{x}_t = x_t\right\},$$

for all $(x_0, x_1, ..., x_t, x_{t+1}) \in S^{t+2}$ and $t = 0, 1, ...$

- The elements of $S$ are called states, $S = \{s_1, s_2, ...\}$.

- A Markov chain is said to be finite if the state space $S$ is finite.

- Usually, the subindex $t$ refers to "time" in Markov processes.
- The Markovian assumption is not so restrictive. To see this, define $\widetilde{X}_t = (\tilde{x}_t, \tilde{x}_{t-1}, ..., \tilde{x}_{t-j})$.
- If $\left\{\widetilde{X}_t\right\}_{t=0}^{\infty}$ is a Markov chain then

$$P\left\{\widetilde{X}_{t+1} = X_{t+1} | \widetilde{X}_t = X_t\right\} =$$

$$P\left\{(\tilde{x}_{t+1}, ..., \tilde{x}_{t-j+1}) = (x_{t+1}, ..., x_{t-j+1}) \,|\, (\tilde{x}_t, ..., \tilde{x}_{t-j}) = (x_t, ..., x_{t-j})\right\}.$$

- Therefore, the conditional distribution of $\tilde{x}_{t+1}$ depends on the values $x_{t-s}$ taken by $\tilde{x}_{t-s}$ for $s = 0, 1, ..., j$.

- Consider from now on a finite Markov chain with $\#S = N < \infty$.
- A transition matrix $\Pi$ is a matrix whose element $\pi_{ij}$ at the $(i, j)$-entry is the probability of $\tilde{x}_{t+1} = s_j$ given $\tilde{x}_t = s_i$,

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1N} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2N} \\ \ldots & \ldots & \ddots & \ldots \\ \pi_{N1} & \pi_{N2} & \cdots & \pi_{NN} \end{pmatrix},$$

where

$$\pi_{ij} = P\left\{\tilde{x}_{t+1} = s_j \middle| \tilde{x}_t = s_i\right\}.$$

- $\pi_{ij}$ is called the transition probability from state $s_i$ to state $s_j$.

- Although one could consider Markov chains whose transition matrices vary with $t$, Markov chains are usually understood to have time-invariant transition matrices unless otherwise specified.

- Any row $i$ of a transition matrix consists of nonnegative elements which must add up to unity, $\sum_j \pi_{ij} = 1$. A matrix with this property is called a **stochastic matrix** and its rows are called **probabilistic (or probability) vectors .**

- The product of transition matrices is also a stochastic matrix that gives the conditional probabilities after a sequence of transitions.

- To see this, observe that, according to the theorem of total (conditional) probability, we have the following:

$$P\left\{\tilde{x}_2 = s_j \middle| \tilde{x}_0 = s_i\right\} = \sum_{m=1}^{N} P\left\{\tilde{x}_2 = s_j, \tilde{x}_1 = s_m \middle| \tilde{x}_0 = s_i\right\}$$

$$= \sum_{m=1}^{N} P\left\{\tilde{x}_1 = s_m \middle| \tilde{x}_0 = s_i\right\} \cdot P\left\{\tilde{x}_2 = s_j \middle| \tilde{x}_1 = s_m, \tilde{x}_0 = s_i\right\}$$

$$= \sum_{m=1}^{N} P\left\{\tilde{x}_1 = s_m \middle| \tilde{x}_0 = s_i\right\} \cdot P\left\{\tilde{x}_2 = s_j \middle| \tilde{x}_1 = s_m\right\} = \sum_{m=1}^{N} \pi_{im} \cdot \pi_{mj},$$

where the first equality follows as $\left\{\tilde{x}_2 = s_j\right\} = \bigcup_{m=1}^{N} \left\{\tilde{x}_2 = s_j, \tilde{x}_1 = s_m\right\}$ and the events appearing in the previous union are disjoint, the second equality is obvious from the definition of conditional probability, and the third equality follows from the fact that $P\left\{\tilde{x}_2 = s_j \middle| \tilde{x}_1 = s_m, \tilde{x}_0 = s_i\right\} = P\left\{\tilde{x}_2 = s_j \middle| \tilde{x}_1 = s_m\right\}$ as dictated by the assumption that $\left\{\tilde{x}_t\right\}_{t=0}^{\infty}$ is a Markov chain.

- Note that $\pi_{ij,2} \equiv \sum_{m=1}^{N} \pi_{im} \cdot \pi_{mj}$ is the value appearing in the $(i,j)$-entry of the square of the time-invariant transition matrix, $\Pi^2$.
- Therefore, in general, the element $\pi_{ij,t}$ in the $(i,j)$-entry of the matrix $\Pi^t$ is

$$\pi_{ij,t} = P\left\{\tilde{x}_t = s_j \middle| \tilde{x}_0 = s_i\right\} = P\left\{\tilde{x}_{n+t} = s_j \middle| \tilde{x}_n = s_i\right\},$$

for all $n = 0, 1, ...$

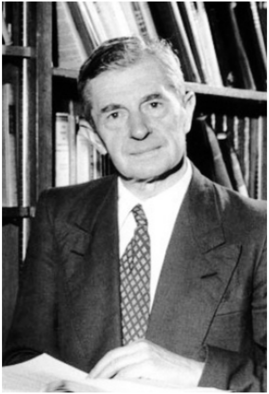
- Chapman-Kolmogorov equation:

$$\pi_{ij,t+s} \equiv \sum_{m=1}^{N} \pi_{im,t} \cdot \pi_{mj,s} \Longleftrightarrow \Pi^{t+s} = \Pi^t \Pi^s.$$

- Let the probabilistic row vector $p_0 = (p_{01}, p_{02}, ..., p_{0N})$, where $p_{0i} = P\{\tilde{x}_0 = s_i\}$ with $s_i \in S$, correspond to the initial distribution of states. A Markov chain with the transition matrix $\Pi$ satisfies

$$P\{\tilde{x}_0 = s_{i_0}, \tilde{x}_1 = s_{i_1}, \tilde{x}_2 = s_{i_2}, ..., \tilde{x}_T = s_{i_T}\} = p_{0i_0} \cdot \pi_{i_0 i_1} \cdot \pi_{i_1 i_2} \cdot ... \cdot \pi_{i_{T-1} i_T},$$

for $T = 1, 2, ...$ and all $(s_{i_0}, s_{i_1}, ..., s_{i_T}) \in S^{T+1}$.



Sydney Chapman (1888 - 1970)

- After $t$ transitions, the unconditional probability to be in the state $s_j$ is the $j$th coordinate of a row vector $p_t$, where

$$p_t = p_0 \Pi^t.$$

- Clearly, $p_t$ evolves according to

$$p_{t+1} = p_t \Pi.$$

- The stationary distribution of a Markov chain is characterized by a probabilistic row vector $p^*$ such that

$$p^* = p^* \Pi,$$

  and, thus,

$$p^* = p^* \Pi^t, \text{ for all } t.$$

- $p^*$ is called the **stationary probabilistic vector** and its components are called the **stationary probabilities.**

- **Definition.** Given a square matrix $A$ with $n$ rows/columns, a non-zero column vector $x \in \mathbb{R}^n$ is a (right-)**eigenvector** of $A$ if it satisfies the equation $Ax = \lambda x$ for some scalar $\lambda \in \mathbb{R}$. In this case, the scalar $\lambda$ is the **eigenvalue** of $A$ associated with the (right-)eigenvector $x$, or the vector $x$ is a (right-)eigenvector associated with the eigenvalue $\lambda$.

- Note that the equation $Ax = \lambda x$ can be written as $Ax = \lambda \mathrm{I} x$ or as $(A - \lambda \mathrm{I})x = \underline{0}$, where $\underline{0}$ is a column vector of zeroes and $\mathrm{I}$ is the identity matrix.

- Therefore, since $x$ is a non-zero vector, an eigenvalue is a solution (real or complex) to the equation

$$\det(A - \lambda \mathrm{I}) = 0,$$

and, hence, the set of vectors $x$ satisfying the equation $(A - \lambda \mathrm{I})x = \underline{0}$, where $\lambda$ is an eigenvalue of $A$, constitutes a vector subspace of $\mathbb{R}^n$.

- **Definition.** The (right-)**eigenspace** associated with a particular eigenvalue of a matrix is the set of the (right-)eigenvectors associated with this eigenvalue, together with the zero vector (which has no direction). That is, the (right-)eigenspace associated with the eigenvalue $\lambda$ of the matrix $A$ is the set of vectors $x$ satisfying the equation $(A - \lambda \mathrm{I})x = \underline{0}$.

- *Note:* If $x$ is a (right-)eigenvector associated with the eigenvalue $\lambda$, then the vector $\alpha x$, where $\alpha \in \mathbb{R}$ is an arbitrary scalar, belongs to the (right-)eigenspace associated with the eigenvalue $\lambda$. Moreover, any linear combination of (right-)eigenvectors associated with the eigenvalue $\lambda$ belongs to the (right-)eigenspace associated with the eigenvalue $\lambda$.

- **Definition.** Given a square matrix $A$ with $n$ rows/columns, a non-zero row vector $x \in \mathbb{R}^n$ is a **left-eigenvector** of $A$ if it satisfies the equation $xA = \lambda x$ for some scalar $\lambda \in \mathbb{R}$. In this case, the vector $x$ is the left-eigenvector associated with the "left-eigenvalue" $\lambda$.

- *Note:* A scalar $\lambda$ satisfying the equation $xA = \lambda x$ (or $xA = \lambda x\mathrm{I}$ or $x(A - \lambda \mathrm{I}) = \underline{0}^\intercal$) for some non-zero row vector $x$ must be also a solution to the equation $\det(A - \lambda \mathrm{I}) = 0$. That is, the set of "left-eigenvalues" coincides with the set of eigenvalues. However, the sets of (transposed) left-eigenvectors and of (right-)eigenvectors are not equal.

- **Definition.** The **left-eigenspace** associated with a particular eigenvalue of a matrix is the set of the left-eigenvectors associated with this eigenvalue, together with the zero vector (which has no direction). That is, the left-eigenspace associated with the eigenvalue $\lambda$ of the matrix $A$ is the set of row vectors $x$ satisfying the equation $x(A - \lambda \mathrm{I}) = \underline{0}^\intercal$.

- A stochastic matrix has always an eigenvalue equal to 1 and the column vector $\underline{1} \equiv (1, \ldots, 1)^\intercal$ is one of the (right-)eigenvectors associated with the eigenvalue 1. Obviously, since

$$\sum_j \pi_{ij} = 1, \text{ for all } i \iff \Pi \underline{1} = \underline{1} \iff (\Pi - \mathrm{I})\, \underline{1} = \underline{0},$$

we must have

$$\det (\Pi - \mathrm{I}) = 0.$$

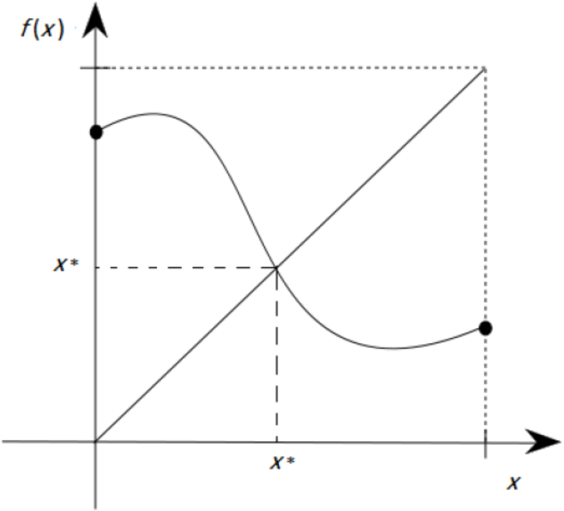
- Therefore, any column vector with identical values in its entries, $\underline{\alpha} = \alpha \underline{1} = (\alpha, \alpha, \ldots, \alpha)^\intercal$, belongs to the (right-)eigenspace associated with the eigenvalue 1 of the stochastic matrix $\Pi$.

- The stationary probabilistic vector $p^*$ of the Markov chain is a left-eigenvector associated with the eigenvalue 1 since

$$p^* = p^* \Pi \iff p^* (\Pi - \mathrm{I}) = \underline{0}^\intercal,$$

where $\underline{0}^\intercal$ is a row vector of zeroes ($\underline{0}$ is a column vector of zeroes). Moreover, the elements of the vector $p^*$ must be non-negative and add up to unity.

- **Brouwer's fixed point theorem:** Every continuous function $f$ from a convex compact subset $D$ of $\mathbb{R}^n$ to itself has a fixed point $x^*$, $f(x^*) = x^*$.



- Consider the unitary simplex $\Delta$ of $\mathbb{R}^n$,

$$\Delta = \left\{ (x_1, x_2, ..., x_n) \in \mathbb{R}^n \;\middle|\; x_i \in [0,1]\,,\ \text{for } i = 1, 2, ..., n,\ \text{and } \sum_{i=1}^{n} x_i = 1 \right\}.$$

  - *Note:* The dimension of a set in $\mathbb{R}^n$ is the minimum number of coordinates needed to specify any point within it.
  - Hence, the unitary simplex has dimension $n - 1$.
  - The set of probabilistic vectors is the unitary simplex $\Delta$.
  - The unitary simplex $\Delta$ is convex and compact and the function $f : \Delta \longrightarrow \Delta$ defined by $f(p) = p\Pi$, where $\Pi$ is a stochastic matrix, is continuous. Therefore, according to Brouwer's fixed point theorem, $f$ has a fixed point $p^*$, $f(p^*) = p^*$, that is, there exists a probabilistic vector $p^*$ such that $p^*\Pi = p^*$.

- Therefore, the stationary probabilistic vector $p^*$ always exists but may fail to be unique, since the eigenvalue 1 could exhibit a multiplicity.
- The dimension of the (non-empty) set of stationary probabilistic vectors is equal to the multiplicity of the eigenvalue 1 of the matrix $\Pi$ minus 1.
- If there exists a matrix $\widehat{\Pi}$ such that $\lim_{t\to\infty} \Pi^t = \widehat{\Pi}$, then the row $i$ of the limiting matrix $\widehat{\Pi}$ gives the conditional long-run (or ergodic) probabilities of the different states when the initial state is $s_i$. These conditional ergodic probabilities define the "conditional ergodic distribution" of the chain given $\tilde{x}_0 = s_i$.
- The element $\widehat{\pi}_{ij}$ at row $i$ and column $j$ of the matrix $\widehat{\Pi}$ is thus

$$\widehat{\pi}_{ij} = \lim_{t\to\infty} P\left\{\tilde{x}_t = s_j \middle| \tilde{x}_0 = s_i\right\}.$$

- Moreover, if all the rows of $\widehat{\Pi}$ are identical and equal to the row vector $\widehat{p}$, then this probabilistic vector $\widehat{p}$ gives the "ergodic distribution" of the Markov chain, which is the long-run distribution of the Markov chain for all initial probabilistic vectors $p_0$. The row vector $\widehat{p}$ is called the **ergodic probabilistic vector** of the Markov chain and its components are called the **ergodic probabilities.** Clearly, in this case,

$$\lim_{t\to\infty} p_0 \Pi^t = p_0 \widehat{\Pi} = \widehat{p}, \text{ for all } p_0.$$

- Moreover, as

$$p^* = \lim_{t\to\infty} p^* \Pi^t = p^* \widehat{\Pi} = \widehat{p},$$

  the ergodic distribution, if it exists, coincides with the stationary distribution, which is unique in this case.

- However, $\lim_{t\to\infty} \Pi^t$ may fail to exist. Moreover, even if this limit exists, it may occur that the limiting matrix $\widehat{\Pi}$ has non-identical rows. In these cases, there is no ergodic probabilistic vector and, hence, the ergodic distribution does not exist.

- Recall that a stochastic matrix has always an eigenvalue equal to 1 and that the column vector $\underline{1} \equiv (1, ..., 1)^{\intercal}$ is one of the (right-)eigenvectors associated with the eigenvalue 1. Therefore, any column vector with identical values in its entries, $\underline{\alpha} = \alpha \underline{1} = (\alpha, \alpha, ..., \alpha)^{\intercal}$, belongs to the (right-)eigenspace associated with the eigenvalue 1.

- While stochastic matrices have at least one eigenvalue equal to 1, the other (real or complex) eigenvalues have modulus smaller or equal than 1.

- A sufficient condition for the existence of the ergodic distribution is that the transition matrix $\Pi$ be **regular** (i.e., there exists a natural number $t$ for which the matrix $\Pi^t$ has only strictly positive elements). In this case, $\widehat{p} = p^* \gg 0$.

- **Doeblin's Theorem:** A Markov chain has an ergodic distribution if and only if there is a natural number $t$ for which $\Pi^t$ has a column $j$ whose elements are all strictly positive. In this case, $\widehat{p}_j = p_j^* > 0$.

- Note that, if the column $j$ has all its elements strictly positive, then the state $s_j$ has always a strictly positive probability of being reached for all the states of the chain in the previous period. The state $s_j$ is going to visited again and again and, after a while, the chain's initial distribution is going to get "forgotten."

- Please, read the Introduction and Sections 1 and 2 of Bru, B. and Yor, M. (2002). "Comments on the life and mathematical legacy of Wolfgang Doeblin." *Finance and Stochastics* 6: 3-47. https://link.springer.com/content/pdf/10.1007%2Fs780-002-8399-0.pdf

- An equivalent necessary and sufficient condition for the existence of the ergodic distribution is that the matrix $\Pi$ has a unique eigenvalue equal to 1 and all the other eigenvalues have modulus strictly smaller than 1.

Wolfgang Doeblin (1915 – 1940)

- If the limiting matrix $\widehat{\Pi}$ exists, then

$$\Pi \cdot \widehat{\Pi} = \Pi \cdot \left( \lim_{t \to \infty} \Pi^t \right) = \lim_{t \to \infty} \Pi^{t+1} = \widehat{\Pi}.$$

- Therefore, if the column vector $\overline{\pi}_j$ is the column $j$ of $\widehat{\Pi}$, then $\Pi \overline{\pi}_j = \overline{\pi}_j$ (or $(\Pi - \mathrm{I})\, \overline{\pi}_j = \underline{0}$) so that $\overline{\pi}_j$ belongs to the (right-)eigenspace associated with the eigenvalue 1 of the matrix $\Pi$.

- In fact, the column vectors of $\widehat{\Pi}$ span the (right-)eigenspace associated with the eigenvalue 1 of the matrix $\Pi$.

- Thus, if the matrix $\Pi$ has only one eigenvalue equal to 1 and all the other eigenvalues have modulus strictly smaller than 1, then the limiting matrix $\widehat{\Pi}$ exists and all the columns of $\widehat{\Pi}$ will be proportional to a column vector $\underline{\alpha} = (\alpha, \alpha, ..., \alpha)^{\intercal}$, with identical values in its entries, which implies in turn that all the rows of $\widehat{\Pi}$ are identical and, hence, the ergodic distribution exists. Moreover, in this case the stationary probabilistic vector is unique. **(Case I)**

- If the matrix $\Pi$ has several eigenvalues equal to 1 and all the other eigenvalues have modulus strictly smaller than one, then the limiting matrix $\widehat{\Pi}$ exists but its rows are non-identical. Therefore, the ergodic distribution does not exist. Moreover, in this case the set of stationary probabilistic vectors has dimension equal to the multiplicity of the eigenvalue 1 of the matrix $\Pi$ minus 1. **(Case II)**
- If the matrix $\Pi$ has some eigenvalues different from 1 but with modulus equal to 1, then the limiting matrix $\widehat{\Pi}$ fails to exist. Therefore, the ergodic distribution does not exist. Moreover, in this case the set of stationary probabilistic vectors has also dimension equal to the multiplicity of the eigenvalue 1 of the matrix $\Pi$ minus 1. **(Case III)**

- Moreover, if the limiting matrix $\widehat{\Pi}$ exists, then

$$\widehat{\Pi} \cdot \Pi = \left( \lim_{t \to \infty} \Pi^t \right) \cdot \Pi = \lim_{t \to \infty} \Pi^{t+1} = \widehat{\Pi}.$$

- Therefore, if the row vector $\overline{\pi}_i$ is the row $i$ of $\widehat{\Pi}$, then $\overline{\pi}_i \Pi = \overline{\pi}_i$ (or $\overline{\pi}_i (\Pi - \mathrm{I}) = \underline{0}^\intercal$) so that $\overline{\pi}_i$ belongs to the left-eigenspace associated with the eigenvalue 1 of the matrix $\Pi$.

- In fact, the row vectors of $\widehat{\Pi}$ span the left-eigenspace associated with the eigenvalue 1 of the matrix $\Pi$.

- Recall also that the row $i$ of the limiting matrix $\widehat{\Pi}$ gives the conditional long-run (or ergodic) probabilities of the different states when the initial state is $s_i$. These conditional ergodic probabilities define the "conditional ergodic distribution" of the Markov chain given $\tilde{x}_0 = s_i$.

- **Definition.** A state $s_i \in S$ is absorbing if

$$\pi_{ii} = P\left\{\tilde{x}_{t+1} = s_i \middle| \tilde{x}_t = s_i\right\} = 1.$$

- If the state $s_i$ is absorbing then a probabilistic vector that assigns probability 1 to the state $s_i$ (and zero probability to the other states) is a stationary probabilistic vector.

- Assume that the limiting matrix $\widehat{\Pi}$ exists. If the state $s_j$ is absorbing then the $(i,j)$-entry $\widehat{\pi}_{ij}$ of the matrix $\widehat{\Pi}$ gives the probability that a chain starting in state $s_i$ ends up being absorbed by state $s_j$.

- **Definition.** A set $E$ is ergodic if

  (a) $E$ is a subset of the state space $S$.

  (b) $P\left\{\tilde{x}_{t+1} \in E \middle| \tilde{x}_t \in E\right\} = 1.$

  (c) No proper (or strict) subset of $E$ satisfies property (b).

- *Note:* $D$ is a proper (or strict) subset of $E$ if $D \subset E$ and $D \neq E$.

- *Note:* Absorbing states are ergodic sets with $\#E = 1$.

- **Definition.** A state $s_i \in S$ is transient if, given $\tilde{x}_0 = s_i$, there is a non-zero probability that we will never return to $s_i$, i.e., if

$$P\left( \bigcap_{t=1}^{\infty} \{\tilde{x}_t \neq s_i\} \middle| \tilde{x}_0 = s_i \right) > 0.$$

- **Definition.** A state $s_i \in S$ is recurrent (or persistent) if it is not transient, i.e., if

$$P\left( \bigcap_{t=1}^{\infty} \{\tilde{x}_t \neq s_i\} \middle| \tilde{x}_0 = s_i \right) = 0$$

or

$$P\left( \bigcup_{t=1}^{\infty} \{\tilde{x}_t = s_i\} \middle| \tilde{x}_0 = s_i \right) = 1.$$

- **Definition.** A state $s_i \in S$ has period $k$ if any return to state $s_i$ must occur in multiples of $k$ transitions (or steps or time periods).
- Formally, the period of a state $s_i$ is defined as the greatest common divisor of $\{t \mid P\{\tilde{x}_t = s_i | \tilde{x}_0 = s_i\} > 0\}$, where $t$ is a strictly positive natural number.
- Note that even though a state has period $k$, it may not be possible to reach the state in $k$ steps. For example, suppose it is possible to return to the state in $\{6, 9, 12, 15, ...\}$ steps; $k$ would be 3, even though 3 does not appear in this list.
- If $k = 1$, then the state $s_i$ is said to be aperiodic: returns to state $s_i$ can occur at irregular times. Otherwise $(k > 1)$, the state $s_i$ is said to be periodic with period $k$. Thus, absorbing states are aperiodic.

- **Examples.** The state space is $S = \{s_1, s_2, s_3\}$.
- **Case I.**

$$\lim_{t\to\infty} \underbrace{\begin{pmatrix} 3/10 & 0 & 7/10 \\ 4/5 & 1/5 & 0 \\ 0 & 3/5 & 2/5 \end{pmatrix}}_{\Pi}^{t} = \begin{pmatrix} 24/73 & 21/73 & 28/73 \\ 24/73 & 21/73 & 28/73 \\ 24/73 & 21/73 & 28/73 \end{pmatrix} = \widehat{\Pi}$$

- None of the columns of the transition matrix $\Pi$ has all its elements strictly positive. However, the eigenvalues of the matrix $\Pi$ are 1 and $-0.05 \pm 0.59791i$ (which have modulus smaller than 1) so that the ergodic distribution exists. In fact, the matrix $\Pi$ is regular. In this case,

$$p^* = \widehat{p} = \left(\frac{24}{73}, \frac{21}{73}, \frac{28}{73}\right)$$

and the single ergodic set is the state space $S = \{s_1, s_2, s_3\}$. All the states are recurrent (or persistent) and aperiodic.

• **Case II.**

$$\lim_{t\to\infty}\underbrace{\begin{pmatrix} 2/5 & 1/5 & 2/5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{\Pi}^{t} = \begin{pmatrix} 0 & 1/3 & 2/3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \widehat{\Pi}$$

• The matrix $\widehat{\Pi}$ exists but has non-identical rows. The eigenvalues of the transition matrix $\Pi$ are $1, 1$ and $0.4$. Hence, there is no ergodic probabilistic vector and the ergodic distribution fails to exist. Moreover, all the columns of $\widehat{\Pi}$ belong to (and they span) the (right-)eigenspace associated with the eigenvalue $1$ of $\Pi$.

• The Markov chain has the following stationary probabilistic vectors:

$$p^* = (0,\ \beta,\ 1-\beta) \text{ for all } \beta \in [0,1],$$

which form a subset of $\mathbb{R}^3$ of dimension $1$ $(= 2-1)$. Clearly, $(0, \beta, 1-\beta)\cdot\Pi = (0, \beta, 1-\beta)$.

• The states $s_2$ and $s_3$ are absorbing and there are two ergodic sets: $\{s_2\}$ and $\{s_3\}$. The state $s_1$ is transient whereas the states $s_2$ and $s_3$ are obviously recurrent (or persistent). All three states are aperiodic.

- **Case III.**

$$\lim_{t\to\infty}\underbrace{\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 3/10 & 2/10 & 1/2 \end{pmatrix}^t}_{\Pi}$$

does not exists since

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 3/10 & 2/10 & 1/2 \end{pmatrix}^t = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ a_t & b_t & c_t \end{pmatrix}$$

if $t$ is odd, whereas

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 3/10 & 2/10 & 1/2 \end{pmatrix}^t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ d_t & e_t & f_t \end{pmatrix}$$

if $t$ is even. Thus, there is no ergodic distribution.

- The eigenvalues of the transition matrix $\Pi$ are $1, -1$ and $0.5$. Thus, since there is an eigenvalue different from 1 with modulus equal to 1, $\lim_{t \to \infty} \Pi^t$ does not exist and there is no ergodic probabilistic vector.

- Since there is only one eigenvalue equal to 1, the Markov chain has the following unique stationary probabilistic vector:

$$p^* = \left( \frac{1}{2}, \ \frac{1}{2}, \ 0 \right).$$

- The single ergodic set is $\{s_1, s_2\}$. The state $s_3$ is transient and aperiodic. The states $s_1$ and $s_2$ are recurrent (or persistent) and periodic with period 2.

- **A $4 \times 4$ matrix example (Case II) - Exercise**
- The matrix

$$\Pi' = \begin{pmatrix} 1/5 & 0 & 4/5 & 0 \\ 0 & 3/5 & 0 & 2/5 \\ 7/10 & 0 & 3/10 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix}$$

could be rewritten, without loss of generality, as

$$\Pi = \begin{pmatrix} 1/5 & 4/5 & 0 & 0 \\ 7/10 & 3/10 & 0 & 0 \\ 0 & 0 & 3/5 & 2/5 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

by permuting or relabeling the states $s_2$ and $s_3$. In this way, the two ergodic sets become more apparent.

$$\lim_{t\to\infty} \underbrace{\begin{pmatrix} 1/5 & 4/5 & 0 & 0 \\ 7/10 & 3/10 & 0 & 0 \\ 0 & 0 & 3/10 & 4/10 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}^t}_{\Pi} = \underbrace{\begin{pmatrix} 7/15 & 8/15 & 0 & 0 \\ 7/15 & 8/15 & 0 & 0 \\ 0 & 0 & 5/9 & 4/9 \\ 0 & 0 & 5/9 & 4/9 \end{pmatrix}}_{\widehat{\Pi}}$$

- The matrix $\widehat{\Pi}$ exists but has non-identical rows. The eigenvalues of the transition matrix $\Pi$ are $1, 1, 0.1$, and $-0.5$. Hence, there is no ergodic probabilistic vector and the ergodic distribution fails to exist. Moreover, all the columns of $\widehat{\Pi}$ belong to (and they span) the (right-)eigenspace associated with the eigenvalue 1 of $\Pi$.

- The Markov chain has the following stationary probabilistic vectors:

$$p^* = \left(\alpha,\ \frac{8\alpha}{7},\ \frac{5}{9} - \frac{25\alpha}{21},\ \frac{4}{9} - \frac{20\alpha}{21}\right) \quad \text{for all } \alpha \in [0, 7/15],$$

which form a subset of $\mathbb{R}^4$ of dimension 1 $(= 2 - 1)$. Clearly,

$$\left(\alpha,\ \frac{8\alpha}{7},\ \frac{5}{9} - \frac{25\alpha}{21},\ \frac{4}{9} - \frac{20\alpha}{21}\right) \cdot \Pi = \left(\alpha,\ \frac{8\alpha}{7},\ \frac{5}{9} - \frac{25\alpha}{21},\ \frac{4}{9} - \frac{20\alpha}{21}\right).$$

- If $\alpha = 7/15$, then we remain all the time in the (relabeled) ergodic set $\{s_1, s_2\}$. If $\alpha = 0$, then we remain all the time in the (relabeled) ergodic set $\{s_3, s_4\}$.

- There are no absorbing states and there are two (relabeled) ergodic sets: $\{s_1, s_2\}$ and $\{s_3, s_4\}$. All the states are recurrent and aperiodic.

- **Absorbing Markov Chains:**
- **Definition.** A Markov chain is **absorbing** if it has at least one absorbing state and if from every state it is possible to go to an absorbing state (not necessarily in one step).
- **Questions:**
  1. What is the probability that the chain will eventually reach an absorbing state?
  2. On the average, how many times will the chain be in each transient state?
  3. On the average, how long will it take for the chain to be absorbed?
  4. What is the probability that the chain will end up in a given absorbing state?
- See the handout for the answers to these four questions.

- **Irreducible Markov Chains:**

- **Definition.** A Markov chain is called **irreducible** (also called **ergodic**) if it is possible to go from every state to every state (not necessarily in one step).

- **Questions:**

  1. On the average, how long will it take for the chain to reach state $s_j$ for the first time if it started in state $s_i$?
  2. If the chain is started in state $s_i$, how long will it take on average to return to $s_i$ for the first time?

- See the handout for the answers to these two questions.

- **General Markov processes:**
- **Definition.** A Markov process (or chain) is a sequence of random objects $\{\tilde{x}_t\}_{t=0}^{\infty}$ , $\tilde{x}_t : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega', \mathcal{F}')$ , where the conditional distribution of $\tilde{x}_{t+1}$ given $(\tilde{x}_t, \tilde{x}_{t-1}, ..., \tilde{x}_0) = (x_t, x_{t-1}, ..., x_0)$ depends only on the value $x_t$ taken by $\tilde{x}_t$ for $t = 0, 1, ...,$

$$P\left\{\tilde{x}_{t+1} \in B \middle| \tilde{x}_t = x_t, \tilde{x}_{t-1} = x_{t-1}, ..., \tilde{x}_0 = x_0\right\}$$

$$= P\left\{\tilde{x}_{t+1} \in B \middle| \tilde{x}_t = x_t\right\},$$

for all $B \in \mathcal{F}'$, all $(x_0, x_1, ..., x_t) \in \left(\Omega'\right)^{t+1}$, and $t = 0, 1, ...$

- A Markov process $\{\tilde{x}_t\}_{t=0}^{\infty}$ has time-invariant transition if there exists a time-invariant function called the **transition function** $Q : \Omega' \times \mathcal{F}' \longrightarrow [0,1]$, which is given by

$$Q(x,B) = P\{\tilde{x}_{t+1} \in B \,|\tilde{x}_t = x\}, \text{ for all } x \in \Omega', \ B \in \mathcal{F}', \ t = 0,1,...$$

so that the distribution $P_{t+1}$ of the random object $\tilde{x}_{t+1}$ is given by the following recursive formula (which arises from the theorem of total probability):

$$P_{t+1}(B) = \int_{\Omega'} Q(x,B)dP_t(x), \text{ for all } B \in \mathcal{F}', \ t = 0,1,...$$

- Thus, the stationary distribution $P^*$ of the Markov process satisfies

$$P^*(B) = \int_{\Omega'} Q(x,B)dP^*(x), \text{ for all } B \in \mathcal{F}'.$$

- Assume now that the random objects $\{\tilde{x}_t\}_{t=0}^{\infty}$ are random variables, $\tilde{x}_t : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$.
- The Markov process of random variables $\{\tilde{x}_t\}_{t=0}^{\infty}$ has an ergodic distribution $\widehat{P}$ if $P_t \xrightarrow{w} \widehat{P}$ for all $P_0$. In this case, $\widehat{P} = P^*$ and the stationary distribution $P^*$ is unique.
- Thus, a Markov process of random variables $\{\tilde{x}_t\}_{t=0}^{\infty}$ having the ergodic distribution $\widehat{P}$ converges in distribution to a random variable $\tilde{x}$ having the distribution $\widehat{P}$, $\tilde{x}_t \xrightarrow{d} \tilde{x}$ or $\tilde{x}_t \longrightarrow \widehat{P}$.

- If the random objects $\{\tilde{x}_t\}_{t=0}^{\infty}$ are absolutely continuous random variables, $\tilde{x}_t : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$, with the density $f_t$, and there exists a time-invariant conditional density $f_{\tilde{x}'|\tilde{x}}(\cdot\,|x) : \mathbb{R} \longrightarrow \overline{\mathbb{R}}$ such that

$$P\{\tilde{x}_{t+1} \in B\,|\tilde{x}_t = x\} = \int_B f_{\tilde{x}'|\tilde{x}}(x'\,|x)dx',$$

$$\text{for all } x \in \mathbb{R},\ B \in \mathcal{B},\ t = 0, 1, ...,$$

then

$$P_{t+1}(B) = \int_B f_{t+1}(x')dx' = \int_B \int_{\mathbb{R}} f_{\tilde{x}_{t+1},\tilde{x}_t}(x', x)dxdx'$$

$$= \int_B \underbrace{\int_{\mathbb{R}} f_{\tilde{x}'|\tilde{x}}(x'\,|x)f_t(x)dx}_{f_{t+1}(x')}dx', \quad \text{for all } B \in \mathcal{B},\ t = 0, 1, ...$$

- Therefore, the densities evolve according to

$$f_{t+1}(x') = \int_{\mathbb{R}} f_{\tilde{x}'|\tilde{x}}(x' \,|x\,) f_t(x) dx, \quad \text{for all } x' \in \mathbb{R}.$$

- Thus, the stationary density $f^*$ of the Markov process satisfies

$$f^*(x') = \int_{\mathbb{R}} f_{\tilde{x}'|\tilde{x}}(x' \,|x\,) f^*(x) dx, \quad \text{for all } x' \in \mathbb{R}.$$

# 6.6. The Poisson distribution as the limit of binomial distributions

Siméon Denis Poisson (1781 - 1840)

- See the handout for the proof of the following statements:
- Let

$$b(x; n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, ..., n.$$

be the probability function of a binomial random variable with the parameters $n$ and $\theta$. Then

$$\lim_{\substack{n \to \infty \\ n\theta = \lambda \in (0,\infty)}} b(x; n, \theta) = \lim_{\substack{\theta \to 0 \\ n\theta = \lambda \in (0,\infty)}} b(x; n, \theta) = \frac{\lambda^x e^{-\lambda}}{x!}, \text{ for } x = 0, 1, \ldots$$

- The random variable $\tilde{x}$ has a Poisson distribution if its probability function is

$$p(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \text{with } \lambda > 0, \quad \text{for } x = 0, 1, \ldots$$

- We write $\tilde{x} \sim \mathrm{P}(\lambda)$.
- Thus, if $\tilde{x}_n \sim \mathrm{B}(n, \theta)$, then $\tilde{x}_n \to \mathrm{P}(\lambda)$ when $n \to \infty$ and $n\theta = \lambda \in (0, \infty)$ (or when $\theta \to 0$ and $n\theta = \lambda \in (0, \infty)$).
- Mean and variance of the Poisson distribution:

$$\mu = n\theta = \lambda,$$

and

$$\sigma^2 = n\theta(1 - \theta) = \lambda,$$

since $n\theta = \lambda$ and $\theta \to 0$.

- Moment generating function of the Poisson distribution, $M_{\tilde{x}}(t)$ :

$$M_{\tilde{x}}(t) = e^{\lambda(e^t - 1)}.$$

- The Poisson approximation to the binomial distribution is good when $n \geq 30$ and $\theta \leq 0.1$.

- **Example.** Let $n = 500$ and $\theta = 0.01$. Thus, $\lambda = n\theta = 5$. Find $P\{\tilde{x} = 7\}$.

- $$P\{\tilde{x} = 7\} \approx p(7;5) = \frac{5^7 e^{-5}}{7!} = 0.10444,$$

- whereas

$$b(7;500,0.01) = \binom{500}{7} (0.01)^7 (0.99)^{493} = 0.10476.$$

Thus, the approximation error is 0.00032.

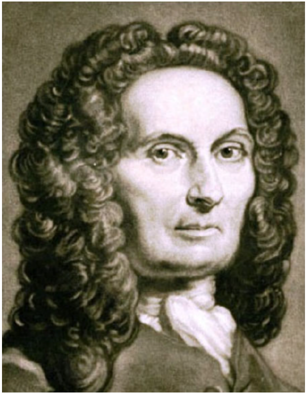
# 6.7. The standard normal distribution as the limit of standardized binomial distributions

- **Theorem (De Moivre-Laplace).** If $\tilde{x}_n$ is a random variable having a binomial distribution with parameters $n$ and $\theta \in (0,1)$, i.e., its probability function is $b(x; n, \theta)$, then the moment generating function of the standardization of $\tilde{x}_n$,

$$\tilde{z}_n = \frac{\tilde{x}_n - \mathrm{E}(\tilde{x}_n)}{\sqrt{\mathrm{Var}(\tilde{x}_n)}} = \frac{\tilde{x}_n - n\theta}{\sqrt{n\theta(1-\theta)}} ,$$

  tends to that of the standard normal distribution as $n \longrightarrow \infty$. Hence, $\tilde{z}_n \longrightarrow \mathrm{N}(0,1)$.

- **Proof.** See the handout.



Abraham de Moivre (1667 - 1754)

- Let $\tilde{x} \sim \mathrm{B}(n, \theta)$ and $a$ and $b$ are natural numbers smaller or equal than $n$, with $a \leq b$. Then, if $n$ is large,

$$P\{a \leq \tilde{x} \leq b\} = P\left\{a - \frac{1}{2} \leq \tilde{x} \leq b + \frac{1}{2}\right\}$$

$$= P\left\{\frac{a - \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}} \leq \frac{\tilde{x} - n\theta}{\sqrt{n\theta(1-\theta)}} \leq \frac{b + \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}}\right\}$$

$$= P\left\{\frac{a - \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}} \leq \tilde{z} \leq \frac{b + \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}}\right\}$$

$$\approx \mathrm{N}\left(\frac{b + \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}}\right) - \mathrm{N}\left(\frac{a - \frac{1}{2} - n\theta}{\sqrt{n\theta(1-\theta)}}\right),$$

  where $\mathrm{N}(\cdot)$ is the distribution function of a standard normal random variable $\tilde{z}$.

- The $\pm\frac{1}{2}$ continuity correction becomes less important as $n$ increases when $a < b$.

- Moreover,

$$P\{\tilde{x}=x\} \approx \mathrm{N}\left(\frac{x+\frac{1}{2}-n\theta}{\sqrt{n\theta(1-\theta)}}\right)-\mathrm{N}\left(\frac{x-\frac{1}{2}-n\theta}{\sqrt{n\theta(1-\theta)}}\right).$$

- The normal approximation to the binomial distribution is good when $n\theta > 5$ and $n(1-\theta) > 5$.
- **Example.** Let $n = 100$ and $\theta = 0.4$ $(n\theta = 40,\ n\theta(1-\theta) = 24)$. Find $P\{\tilde{x} = 35\}$.
- $$P\{\tilde{x}=35\} \approx \mathrm{N}\left(\frac{35+\frac{1}{2}-40}{\sqrt{24}}\right)-\mathrm{N}\left(\frac{35-\frac{1}{2}-40}{\sqrt{24}}\right)$$
$$= \mathrm{N}(-0.91856) - \mathrm{N}(-1.1227) = 0.04838,$$
- whereas

$$b(35; 100, 0.4) = \binom{100}{35}(0.4)^{35}(0.6)^{65} = 0.04913.$$

Thus, the approximation error is 0.00075.

# 6.8. Weak and strong laws of large numbers

- Let $\{\tilde{x}_i\}_{i=1}^{\infty}$ be a collection of random variables (or stochastic process) defined on the probability space $(\Omega, \mathcal{F}, P)$ . Define the random variable "sum", indexed by $n$, as

$$\tilde{S}_n = \sum_{i=1}^{n} \tilde{x}_i,$$

and the random variable "average", indexed by $n$, as

$$\overline{\mathbf{x}}_n = \frac{\tilde{S}_n}{n} = \frac{\sum_{i=1}^{n} \tilde{x}_i}{n}.$$

- **Definition.** The collection $\{\tilde{x}_i\}_{i=1}^{\infty}$ of random variables satisfies the weak law of large numbers (WLLN) if $\overline{\mathbf{x}}_n - \mathrm{E}(\overline{\mathbf{x}}_n) \xrightarrow{p} 0$, i.e.,

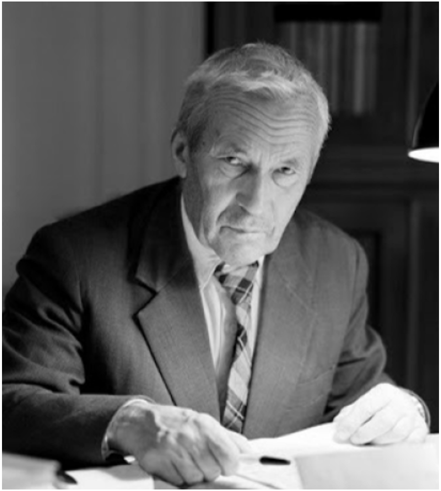
$$\lim_{n\to\infty} P\{|\overline{\mathbf{x}}_n - \mathrm{E}(\overline{\mathbf{x}}_n) - 0| \geq \varepsilon\} = 0 \text{ for all } \varepsilon > 0$$

or

$$\lim_{n\to\infty} P\{|\overline{\mathbf{x}}_n - \mathrm{E}(\overline{\mathbf{x}}_n)| \geq \varepsilon\} = 0, \text{ for all } \varepsilon > 0.$$

- *Note:* $\{\tilde{x}_i\}_{i=1}^{\infty}$ satisfies the WLLN if and only if $\overline{\mathbf{x}}_n - \mathrm{E}(\overline{\mathbf{x}}_n) \xrightarrow{d} 0$. Moreover, if $\overline{\mathbf{x}}_n - \mathrm{E}(\overline{\mathbf{x}}_n) \xrightarrow{m} 0$, then $\{\tilde{x}_i\}_{i=1}^{\infty}$ satisfies the WLLN.
- **Definition.** The collection $\{\tilde{x}_i\}_{i=1}^{\infty}$ of random variables satisfies the strong law of large numbers (SLLN) if $\overline{\mathbf{x}}_n - \mathrm{E}(\overline{\mathbf{x}}_n) \xrightarrow{a.s} 0$, i.e.,

$$P\left\{\lim_{n\to\infty} [\overline{\mathbf{x}}_n - \mathrm{E}(\overline{\mathbf{x}}_n)] = 0\right\} = 1.$$



Andrei Kolmogorov (1903 - 1987)

- **Kolmogorov's Theorem 1.** If

  (a) $\{\tilde{x}_i\}_{i=1}^{\infty}$ are independent random variables with $\tilde{x}_i \in \mathrm{L}^2$ for all $i$, and

  (b)
$$\sum_{i=1}^{\infty} \frac{\mathrm{Var}\left(\tilde{x}_i\right)}{i^2} < \infty,$$

  then $\{\tilde{x}_i\}_{i=1}^{\infty}$ satisfies the SLLN.

- *Note:* Under the assumption of independency,

$$\mathrm{Var}\left(\overline{\mathbf{x}}_n\right) = \mathrm{Var}\left(\frac{\sum_{i=1}^{n} \tilde{x}_i}{n}\right) = \frac{\sum_{i=1}^{n} \mathrm{Var}\left(\tilde{x}_i\right)}{n^2}.$$

- **Particular cases:**

- **(1)** $\{\tilde{x}_i\}_{i=1}^{\infty}$ are independent and $\mathrm{Var}(\tilde{x}_i) \leq M$ for all $i$, where $M < \infty$ is a fixed bound.
  Clearly, in this case,

$$\sum_{i=1}^{\infty} \frac{\mathrm{Var}(\tilde{x}_i)}{i^2} \leq \sum_{i=1}^{\infty} \frac{M}{i^2} = M \cdot \left( \sum_{i=1}^{\infty} \frac{1}{i^2} \right) = M \cdot \frac{\pi^2}{6} < \infty.$$

- **(2)** $\{\tilde{x}_i\}_{i=1}^{\infty}$ are independent random variables with $\mathrm{Var}(\tilde{x}_i) = \sigma^2 < \infty$ for all $i$ (this is a particular case of Case 1).

  In this case,

$$\mathrm{Var}(\overline{\mathbf{x}}_n) = \frac{\sum_{i=1}^{n} \mathrm{Var}(\tilde{x}_i)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

- **(3)** Both (a) and (b) hold and $\mathrm{E}(\tilde{x}_i) = \mu$ for all $i$.

  In this case,

$$\mathrm{E}(\overline{\mathbf{x}}_n) = \mathrm{E}\left(\frac{\sum_{i=1}^{n} \tilde{x}_i}{n}\right) = \frac{\sum_{i=1}^{n} \mathrm{E}(\tilde{x}_i)}{n} = \frac{n\mu}{n} = \mu, \quad \text{for all } n,$$

  so that $\overline{\mathbf{x}}_n - \mu \xrightarrow{a.s} 0$, which can be written as $\overline{\mathbf{x}}_n \xrightarrow{a.s} \mu$ since $P\left\{\lim_{n\to\infty} (\overline{\mathbf{x}}_n - \mu) = 0\right\} = 1$ is equivalent to $P\left\{\lim_{n\to\infty} \overline{\mathbf{x}}_n = \mu\right\} = 1$.

- Similarly, $\overline{\mathbf{x}}_n - \mu \xrightarrow{p} 0$ can be written as $\overline{\mathbf{x}}_n \xrightarrow{p} \mu$.

- *Note:* If $\{\tilde{x}_i\}_{i=1}^{\infty}$ are independent with the finite mean $\mu$ and the finite variance $\sigma^2$ for all $i$, then

$$\mathrm{E}(\overline{\mathbf{x}}_n) = \mu \quad \text{and} \quad \mathrm{Var}(\overline{\mathbf{x}}_n) = \frac{\sigma^2}{n}.$$

- **Kolmogorov's Theorem 2.** If $\{\tilde{x}_i\}_{i=1}^{\infty}$ are i.i.d. random variables with $\tilde{x}_i \in \mathrm{L}^1$ for all $i$, then $\{\tilde{x}_i\}_{i=1}^{\infty}$ satisfies the SLLN.
- *Notes:*

  (a) Theorem 2 implies that $\bar{\mathbf{x}}_n \xrightarrow{a.s} \mu$, where $\mu = \mathrm{E}(\bar{\mathbf{x}}_n) = \mathrm{E}(\tilde{x}_i)$ for all $i$.

  (b) Theorem 1 requires finite variance and $\sum_{i=1}^{\infty} \frac{\mathrm{Var}(\tilde{x}_i)}{i^2} < \infty$ whereas Theorem 2 does not (only finite mean is required in Theorem 2).

  (c) Theorem 2 requires i.i.d. whereas Theorem 1 does not (only independency is required in Theorem 1).

- Let us prove the WLLN for the previous particular case 1, i.e., when $\{\tilde{x}_i\}_{i=1}^{\infty}$ are independent and $\mathrm{Var}(\tilde{x}_i) \leq M$ for all $i$, where $M < \infty$ is a fixed bound.

- First observe that the Chevyshev's inequality,

$$P\left\{\left|\tilde{y} - \mu_{\tilde{y}}\right| \geq k\sigma_{\tilde{y}}\right\} \leq \frac{1}{k^2}, \quad \text{for all } k > 0,$$

can be rewritten as

$$P\left\{\left|\tilde{y} - \mu_{\tilde{y}}\right| \geq \varepsilon\right\} \leq \frac{\sigma_{\tilde{y}}^2}{\varepsilon^2}, \quad \text{for all } \varepsilon > 0.$$

by making the change of variable $k\sigma_{\tilde{y}} = \varepsilon > 0$.

- **Proof.** Make $\tilde{y} = \overline{\mathbf{x}}_n$ so that

$$P\left\{\left|\overline{\mathbf{x}}_n - \mathrm{E}\left(\overline{\mathbf{x}}_n\right)\right| \geq \varepsilon\right\} \leq \frac{\mathrm{Var}\left(\overline{\mathbf{x}}_n\right)}{\varepsilon^2}, \quad \text{for all } \varepsilon > 0.$$

Take limits in both sides,

$$\lim_{n\to\infty} \underbrace{P\left\{\left|\overline{\mathbf{x}}_n - \mathrm{E}\left(\overline{\mathbf{x}}_n\right)\right| \geq \varepsilon\right\}}_{\geq 0} \leq \lim_{n\to\infty} \frac{\mathrm{Var}\left(\overline{\mathbf{x}}_n\right)}{\varepsilon^2}$$

$$= \lim_{n\to\infty} \frac{\sum_{i=1}^{n} \mathrm{Var}\left(\tilde{x}_i\right)}{n^2 \varepsilon^2} \leq \lim_{n\to\infty} \frac{nM}{n^2\varepsilon^2} = \lim_{n\to\infty} \frac{M}{n\varepsilon^2} = 0. \qquad Q.E.D.$$

- **Example:**
- Let $\{\tilde{x}_i\}_{i=1}^{\infty}$ be i.i.d. Bernoulli random variables with parameter $\theta$ so that $\mathrm{E}(\tilde{x}_i) = \theta$ and $\mathrm{Var}(\tilde{x}_i) = \theta(1-\theta)$ for all $i$.
- Then, the random variable $\tilde{S}_n = \sum_{i=1}^{n} \tilde{x}_i$, is binomial, $\tilde{S}_n \sim \mathrm{B}(n, \theta)$.
- Moreover, the random variable

$$\overline{\mathbf{x}}_n = \frac{\tilde{S}_n}{n} = \frac{\sum_{i=1}^{n} \tilde{x}_i}{n}$$

is the percentage (or proportion) of successes in $n$ trials and

$$\mathrm{E}(\overline{\mathbf{x}}_n) = \theta \text{ and } \mathrm{Var}(\overline{\mathbf{x}}_n) = \frac{\theta(1-\theta)}{n}.$$

- Therefore, from any of the two Kolmogorov's theorems, we get

$$\overline{\mathbf{x}}_n \xrightarrow{a.s} \theta,$$

i.e., the percentage (or proportion) of successes converges (almost surely) to the probability of success in each trial when $n \longrightarrow \infty$.

- However, it is not true that the number of successes (almost surely) converges to the expected number of successes since the expected number of successes $\mathrm{E}\left(\tilde{S}_n\right) = n\theta$ tends to infinity as $n \rightarrow \infty$. In fact, even the following is false:

$$\tilde{S}_n - \mathrm{E}\left(\tilde{S}_n\right) \xrightarrow{p} 0 \iff \tilde{S}_n - n\theta \xrightarrow{p} 0. \quad \text{(The fallacy of the LLN)}$$

- Note that, from Chevyshev's theorem, we have

$$P\left(\left|\tilde{S}_n - \mathrm{E}\left(\tilde{S}_n\right)\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}\left(\tilde{S}_n\right)}{\varepsilon^2} = \frac{n\theta(1-\theta)}{\varepsilon^2},$$

and taking limits we get

$$\lim_{n\rightarrow\infty} P\left(\left|\tilde{S}_n - n\theta\right| \geq \varepsilon\right) \leq \lim_{n\rightarrow\infty} \frac{n\theta(1-\theta)}{\varepsilon^2} = \infty,$$

which is a useless inequality!

- **Application:**
- Consider a set of $n$ individuals (or objects). For each individual $i$ there is a random variable $\tilde{z}_i$, $i = 1, ..., n$. The collection of random variables $\{\tilde{z}_i\}_{i=1}^n$ is i.i.d. and $P\{\tilde{z}_i \in A\} = \theta$ for all $i$.
- Let $\tilde{x}_n$ be the number of individuals for whom their respective random variables $\tilde{z}_i$, $i = 1, ..., n$, take a value in the set $A$ ($\tilde{z}_i \in A$).
- Then, the expected number of individuals for which $\tilde{z}_i \in A$, $i = 1, ..., n$, is

$$\mathrm{E}(\tilde{x}_n) = n\theta.$$

- Let $\tilde{y}_n = \dfrac{\tilde{x}_n}{n}$ be the percentage (or proportion) of individuals for which $\tilde{z}_i \in A$, $i = 1, ..., n$.
- Then, the expected percentage (or proportion) of individuals for which $\tilde{z}_i \in A$, $i = 1, ..., n$, is

$$\mathrm{E}(\tilde{y}_n) = \theta.$$

- The law of large numbers tells us not only that $\mathrm{E}(\tilde{y}_n) = \theta$ but also that $\tilde{y}_n \xrightarrow{a.s} \theta$ (and, thus, $\tilde{y}_n \xrightarrow{p} \theta$) as $n \to \infty$.

# 6.9. The Central Limit Theorem (CLT)

- **Theorem (Lindeberg-Lévy).** Let $\{\tilde{x}_i\}_{i=1}^{\infty}$ be i.i.d. random variables with $\mathrm{E}(\tilde{x}_i) = \mu$ and $\mathrm{Var}(\tilde{x}_i) = \sigma^2$, with $0 < \sigma^2 < \infty$ , for all $i$. Then

$$\tilde{z}_n \equiv \frac{\overline{\mathbf{x}}_n - \mathrm{E}(\overline{\mathbf{x}}_n)}{\sqrt{\mathrm{Var}\,(\overline{\mathbf{x}}_n)}} = \frac{\overline{\mathbf{x}}_n - \mu}{\sigma \big/ \sqrt{n}}$$

  converges in distribution to a standard normal random variable as $n \to \infty$. That is, $\tilde{z}_n \longrightarrow \mathrm{N}(0, 1)$.

- **Proof:** See the handout.

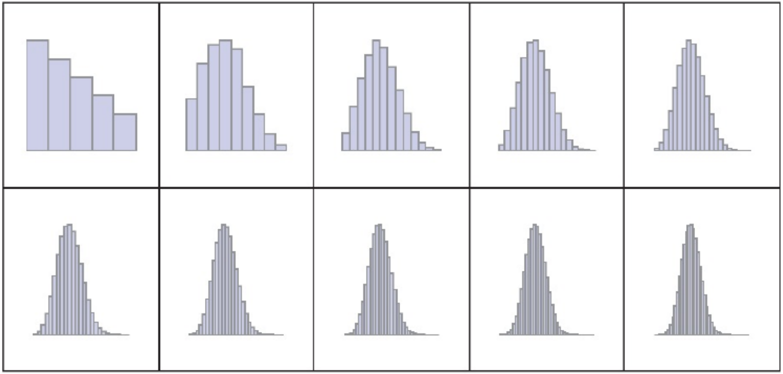
- *Note 1:* The theorem does not say that $\overline{\mathbf{x}}_n \longrightarrow \mathrm{N}\left(\mu, \sigma^2 \big/ n\right)$ since $\sigma^2 \big/ n \longrightarrow 0$ and, thus, $\overline{\mathbf{x}}_n \stackrel{d}{\longrightarrow} \mu$, which is a consequence of any of the two Kolmogrov's theorems.

- *Note 2:* $\tilde{z}_n \equiv \dfrac{\overline{\mathbf{x}}_n - \mu}{\sigma \big/ \sqrt{n}} \longrightarrow \mathrm{N}(0, 1)$ can be written as

$$\sqrt{n}\,(\overline{\mathbf{x}}_n - \mu) \longrightarrow \mathrm{N}(0, \sigma^2).$$

The evolution of the standardization of $\overline{\mathbf{x}}_n$.

| $x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $f_{\tilde{x}}(x)$ | 0.30 | 0.25 | 0.20 | 0.15 | 0.10 |



Jarl W. Lindeberg (1876 – 1932)

- **Other Central Limit Theorems (CLT's):**

- **Theorem (Liapunov).** Let $\{\tilde{x}_i\}_{i=1}^{\infty}$ be independent random variables with finite third moments, $\mathrm{E}(\tilde{x}_i) = \mu_i$, $\mathrm{Var}(\tilde{x}_i) = \sigma_i^2$ with $0 < \sigma_i^2 < \infty$, and $\mathrm{E}[|\tilde{x}_i - \mu_i|^3] = m_{3i}$. If

$$\lim_{j \to \infty} \frac{\left( \sum_{i=1}^{j} m_{3i} \right)^{1/3}}{\left( \sum_{i=1}^{j} \sigma_i^2 \right)^{1/2}} = 0,$$

then $\tilde{z}_n \equiv \dfrac{\overline{\mathbf{x}}_n - \mathrm{E}(\overline{\mathbf{x}}_n)}{\sqrt{\mathrm{Var}\left(\overline{\mathbf{x}}_n\right)}} \longrightarrow \mathrm{N}(0, 1).$

Aleksandr Liapunov (1857 – 1918)

- **Theorem (Lindeberg-Feller).** Let $\{\tilde{x}_i\}_{i=1}^{\infty}$ be independent random variables with distributions $P_i$, $\mathrm{E}(\tilde{x}_i) = \mu_i$, and $\mathrm{Var}(\tilde{x}_i) = \sigma_i^2$ with $0 < \sigma_i^2 < \infty$. Define $c_j = \left( \sum_{i=1}^{j} \sigma_i^2 \right)^{1/2}$. If

$$\lim_{j \to \infty} \frac{1}{c_j^2} \sum_{i=1}^{j} \int_{|x-\mu_i|>\varepsilon c_i} (x - \mu_i)^2 dP_i(x) = 0, \quad \text{for every } \varepsilon > 0,$$

then $\tilde{z}_n \equiv \dfrac{\overline{\mathbf{x}}_n - \mathrm{E}(\overline{\mathbf{x}}_n)}{\sqrt{\mathrm{Var}\left(\overline{\mathbf{x}}_n\right)}} \longrightarrow \mathrm{N}(0,1)$.

- *Note:* The theorems of Lindeberg-Lévy and Liapunov are special cases of the Lindeberg-Feller theorem since the assumptions of any of the first two Central Limit Theorems imply those of the Lindeberg-Feller.

William ("Vilim") Feller (1906 – 1970)

- **Multivariate Central Limit Theorem:**

  Let $\{\tilde{X}_i\}_{i=1}^{\infty}$ be a sequence of $k$-dimensional random vectors, $\tilde{X}_i = (\tilde{x}_{i1}, \tilde{x}_{i2}, ..., \tilde{x}_{ik})^{\top}$. If $\{c^{\intercal}\tilde{X}_i\}_{i=1}^{\infty}$ converges in distribution to a normal random variable for every $k$-dimensional vector of scalars, $c = (c_1, c_2, ..., c_k)^{\top}$ with $c \neq \underline{0}$, then the limiting distribution of $\{\tilde{X}_i\}_{i=1}^{\infty}$ is multivariate normal.

- Notice that showing convergence in distribution to a normal random variable for each random variable appearing in the random vector $\tilde{X}_i$ separately (that is, showing that $\{\tilde{x}_{ij}\}_{i=1}^{\infty}$ converges in distribution to a normal random variable for $j = 1, 2, ..., k$) is not sufficient for the convergence in distribution of the sequence $\{\tilde{X}_i\}_{i=1}^{\infty}$ of random vectors to a random vector having the multivariate normal distribution.

# Proof that convergence in probability implies convergence in distribution.

**Step 1.** Let $\widetilde{y}$ and $\widetilde{z}$ be random variables on the same probability space $(\Omega, \mathcal{F}, P)$, $x$ a real number and $\varepsilon > 0$; then

$$P\{\widetilde{z} \leq x\} = P\{\widetilde{z} \leq x, \widetilde{y} \leq x+\varepsilon\} + P\{\widetilde{z} \leq x, \widetilde{y} > x+\varepsilon\}$$

$$\leq P\{\widetilde{z} \leq x, \widetilde{y} \leq x+\varepsilon\} + P\{\widetilde{z} \leq x, \widetilde{y} \geq x+\varepsilon\} \leq P\{\widetilde{y} \leq x+\varepsilon\} + P\{\widetilde{z} \leq x, x \leq \widetilde{y}-\varepsilon\}$$

$$\leq P\{\widetilde{y} \leq x+\varepsilon\} + P\{\widetilde{z}-\widetilde{y} \leq -\varepsilon\} \leq P\{\widetilde{y} \leq x+\varepsilon\} + P\{|\widetilde{z}-\widetilde{y}| \geq \varepsilon\}$$

since

$$P\{\widetilde{z} \leq x, x \leq \widetilde{y}-\varepsilon\} = P\{\widetilde{z} \leq x \leq \widetilde{y}-\varepsilon\} \leq P\{\widetilde{z} \leq \widetilde{y}-\varepsilon\} = P\{\widetilde{z}-\widetilde{y} \leq -\varepsilon\}$$

and

$$P\{|\widetilde{z}-\widetilde{y}| \geq \varepsilon\} = P\{\widetilde{z}-\widetilde{y} \geq \varepsilon\} + P\{\widetilde{z}-\widetilde{y} \leq -\varepsilon\} \geq P\{\widetilde{z}-\widetilde{y} \leq -\varepsilon\}.$$

Summing up,

$$P\{\widetilde{z} \leq x\} \leq P\{\widetilde{y} \leq x+\varepsilon\} + P\{|\widetilde{z}-\widetilde{y}| \geq \varepsilon\}.$$

**Step 2.** Consider now the sequence of random variables $\{\widetilde{x}_n\}_{n=1}^{\infty}$ and the random variable $\widetilde{x}$ on the same probability space $(\Omega, \mathcal{F}, P)$ such that $\widetilde{x}_n \xrightarrow{p} \widetilde{x}$, that is,

$$\lim_{n\to\infty} P\{|\widetilde{x}_n - \widetilde{x}| \geq \varepsilon\} = 0, \quad \text{for all } \varepsilon > 0$$

Using step 1 we obtain, for every $\varepsilon > 0$ and $x \in \mathbb{R}$,

$$P\{\widetilde{x}_n \leq x\} \leq P\{\widetilde{x} \leq x+\varepsilon\} + P\{|\widetilde{x}_n - \widetilde{x}| \geq \varepsilon\}$$

and, equivalently,

$$P\{\widetilde{x} \leq x-\varepsilon\} \leq P\{\widetilde{x}_n \leq x\} + \underbrace{P\{|\widetilde{x} - \widetilde{x}_n| \geq \varepsilon\}}_{=P\{|\widetilde{x}_n - \widetilde{x}| \geq \varepsilon\}}.$$

Combining the previous two inequalities, we have

$$P\{\widetilde{x} \leq x-\varepsilon\} - P\{|\widetilde{x}_n - \widetilde{x}| \geq \varepsilon\} \leq P\{\widetilde{x}_n \leq x\} \leq P\{\widetilde{x} \leq x+\varepsilon\} + P\{|\widetilde{x}_n - \widetilde{x}| \geq \varepsilon\}.$$

Taking the limit when $n \longrightarrow \infty$, and using the fact that $\widetilde{x}_n \xrightarrow{p} \widetilde{x}$, so that $\lim_{n\to\infty} P(|\widetilde{x}_n - \widetilde{x}| \geq \varepsilon) = 0$ for all $\varepsilon > 0$, we obtain

$$P\{\widetilde{x} \leq x-\varepsilon\} \leq \lim_{n\to\infty} P\{\widetilde{x}_n \leq x\} \leq P\{\widetilde{x} \leq x+\varepsilon\},$$

which can be rewritten as

$$F(x-\varepsilon) \leq \lim_{n\to\infty} F_n(x) \leq F(x+\varepsilon),$$

where $F_n$ is the distribution function of $\widetilde{x}_n$ and $F$ is the distribution function of $\widetilde{x}$.

Assume that the distribution function $F$ is continuous at $x$. Then,

$$\lim_{\varepsilon\to 0} F(x-\varepsilon) \leq \lim_{n\to\infty} F_n(x) \leq \lim_{\varepsilon\to 0} F(x+\varepsilon)$$

becomes

$$F(x) \leq \lim_{n\to\infty} F_n(x) \leq F(x),$$

which means that $\lim_{n\to\infty} F_n(x) = F(x)$ at every continuity point $x$ of $F$. Therefore, $\widetilde{x}_n \xrightarrow{d} \widetilde{x}$. *Q.E.D.*

− − − − − − − − − − − − − − − − −

# Proof that convergence in distribution to a constant implies convergence in probability.

Let $\widetilde{x}_n \xrightarrow{d} a$, where $a$ is a constant. For any $\varepsilon > 0$ we have that

$$P\{|\widetilde{x}_n - a| < \varepsilon\} = P\{a-\varepsilon < \widetilde{x}_n < a+\varepsilon\} = \lim_{z\to(a+\varepsilon)^-} F_n(z) - F_n(a-\varepsilon). \quad (*)$$

Since $\widetilde{x}_n \xrightarrow{d} a$, $F_n$ converges pointwise to

$$F(z) = \begin{cases} 0 & \text{for } z < a \\ 1 & \text{for } z \geq a \end{cases}$$

at every continuity point of $F$, i.e., for all $z \neq a$. This implies that $F_n(a-\varepsilon)$ converges to 0 since $a-\varepsilon < a$, while $F_n(z)$ converges to 1 for all $z > a$. Hence, $\lim_{n\to\infty}\left(\lim_{z\to(a+\varepsilon)^-} F_n(z)\right) = 1$ since $a+\varepsilon > a$.

Using the previous facts, and taking the limit of $(*)$ when $n \longrightarrow \infty$, we get

$$\lim_{n\to\infty} P\{|\widetilde{x}_n - a| < \varepsilon\} = 1 - 0 = 1, \quad \text{for all } \varepsilon > 0,$$

so that $\widetilde{x}_n \xrightarrow{p} a$. *Q.E.D.*

## ABSORBING MARKOV CHAINS

**Definition 1.** A Markov chain is **absorbing** if it has at least one absorbing state and if from every state it is possible to go to an absorbing state (not necessarily in one step).

Obviously, in an absorbing Markov chain, a state that is not absorbing is transient. When a Markov chain reaches an absorbing state, we shall say that it is absorbed.

The most obvious questions that can be asked about an absorbing Markov chain are: What is the probability that the chain will eventually reach an absorbing state? On the average, how many times will the chain be in each transient state? On the average, how long will it take for the chain to be absorbed? What is the probability that the chain will end up in a given absorbing state? The answers to all these questions depend, in general, on the state from which the chain starts as well as the transition probabilities. The next Theorems 1, 2, 3 and 4 provide the answers to those four questions.

Consider an arbitrary absorbing Markov chain with the state space $S = \{s_1, ..., s_N\}$. Renumber the $N$ states so that the transient states come first. If there are $q$ transient states and $d$ absorbing states ($q + d = N$), the $N \times N$ transition matrix $\Pi$ of the chain can be written in the following **canonical form**:

$$\Pi = \begin{array}{c} \\ \text{Trans.} \\ \text{Absorb.} \end{array} \begin{array}{c} \begin{array}{cc} \text{Trans.} & \text{Absorb.} \end{array} \\ \begin{pmatrix} Q & D \\ \widehat{0} & \mathrm{I} \end{pmatrix} \end{array}, \tag{1}$$

where I is a $d \times d$ identity matrix, $\widehat{0}$ is a matrix of zeroes (in this case a $d \times q$ matrix), $D$ is a non-zero $q \times d$ matrix and $Q$ is a $q \times q$ matrix. Thus, now the first $q$ states are transient and the last $d$ states are absorbing.

We know that the entry $\pi_{ij,t}$ of the matrix $\Pi^t$ is the probability of being in state $s_j$ after $t$ steps when the chain is started in state $s_i$. A standard matrix algebra argument shows that $\Pi^t$ has the following form:

$$\Pi^t = \begin{array}{c} \\ \text{Trans.} \\ \text{Absorb.} \end{array} \begin{array}{c} \begin{array}{cc} \text{Trans.} & \text{Absorb.} \end{array} \\ \begin{pmatrix} Q^t & * \\ \widehat{0} & \mathrm{I} \end{pmatrix} \end{array},$$

where the asterisk $*$ stands for the $q \times d$ matrix in the upper right-hand corner of $\Pi^t$. This submatrix can be written in terms of $Q$ and $D$, but the expression is not needed at this time. The form of $\Pi^t$ shows that the entries of $Q^t$ give the probabilities for being in each of the transient states after $t$ steps for each possible transient starting state. In our first theorem we will prove that the probability of being in the transient states after $t$ steps approaches zero. Thus, every entry of $Q^t$ must approach zero as $t$ goes to infinity.

**Theorem 1.** In an absorbing Markov chain, the probability that the chain will be absorbed is 1, i.e., $\lim_{t\to\infty} Q^t = \widehat{0}$, where $\widehat{0}$ is now a $q \times q$ matrix of zeroes

**Proof.** From each non-absorbing state $s_i$ it is possible to reach an absorbing state. Let $m_i$ be the minimum number of steps required to reach an absorbing state, starting from $s_i$. Let $\pi_i$ be the probability that, starting from $s_i$, the chain will not reach an absorbing state in $m_i$ steps. Then $\pi_i < 1$. Let $m$ be the largest of the $m_i$ and let $\pi$ be the largest of the $\pi_i$, $i = 1, ..., q$. The probability of not being absorbed in $m$ steps is less than or equal to $\pi$, in $2m$ steps less than or equal to $\pi^2$, etc. Since $\pi < 1$, these probabilities tend to 0. Therefore, the probability of not being absorbed in $t$ steps is monotone decreasing in $t$ and tends to 0 as $t$ goes to infinity. Hence, $\lim_{t\to\infty} Q^t = \widehat{0}$. *Q.E.D.*

The next theorem shows that the $(i, j)$-entry $n_{ij}$ of the $q \times q$ matrix $(\mathrm{I} - Q)^{-1}$ gives the expected number of times that the chain is in the transient state $s_j$ if it is started in the transient state $s_i$:

**Theorem 2.** For an absorbing Markov chain the matrix $\mathrm{I} - Q$ has an inverse and

$$(\mathrm{I} - Q)^{-1} = \mathrm{I} + Q + Q^2 + Q^3 + ... \equiv \sum_{k=0}^{\infty} Q^k. \tag{2}$$

The $(i, j)$-entry $n_{ij}$ of the matrix $(\mathrm{I} - Q)^{-1}$ is the expected number of times the chain is in the transient state $s_j$ given that it starts in the transient state $s_i$. The initial state is counted when $i = j$.

**Proof.** In order to prove that the matrix $\mathrm{I} - Q$ has an inverse, we will use the fact that a square matrix $G$ has an inverse if and only if $Gx = \underline{0}$ implies that $x = \underline{0}$, where $\underline{0}$ is a column vector of zeroes. Let $(\mathrm{I} - Q)\, x = \underline{0}$, that is, $x = Qx$. Then, iterating this we see that $x = Q^t x$ for all $t$. Since $\lim_{t\to\infty} Q^t = \widehat{0}$, we have that $\lim_{t\to\infty} Q^t x = \underline{0}$, which implies that $x = \underline{0}$. Thus, the inverse matrix $(\mathrm{I} - Q)^{-1}$ exists. Note next that

$$(\mathrm{I} - Q)\left(\mathrm{I} + Q + Q^2 + Q^3 + ... + Q^t\right) = \mathrm{I} - Q^{t+1}.$$

Letting $t$ tend to infinity, from Theorem 1 we get

$$(\mathrm{I} - Q)\,(\mathrm{I} + Q + Q^2 + Q^3 + ...) = \mathrm{I} - \lim_{t\to\infty} Q^{t+1} = \mathrm{I} - \widehat{0} = \mathrm{I}$$

so that (2) holds.

Let $s_i$ and $s_j$ be two transient states and assume throughout the remainder of the proof that $i$ and $j$ are fixed. Let $\tilde{x}_t$ be a random variable that equals 1 if the chain, starting from $s_i$, is in state $s_j$ after $t$ steps, and equals 0 otherwise. For each $t$, this random variable depends upon both $i$ and $j$. We have,

$$P\{\tilde{x}_t = 1\} = q_{ij,t}$$

and

$$P\{\tilde{x}_t = 0\} = 1 - q_{ij,t},$$

where $q_{ij,t}$ is the $(i,j)$-entry of the matrix $Q^t$. These equations hold for $t = 0$ since $Q^0 = \mathrm{I}$. Therefore, since $\tilde{x}_t$ is a 0-1 random variable, $\mathrm{E}(\tilde{x}_t) = q_{ij,t}$.

The expected number of times the chain is in state $s_j$ in the first $k$ periods, given that it starts in state $s_i$, is clearly

$$\mathrm{E}(\tilde{x}_0 + \tilde{x}_1 + \tilde{x}_2 + ... + \tilde{x}_k) = \sum_{t=0}^{k} q_{ij,t}.$$

Letting $k$ tend to infinity we have,

$$\mathrm{E}(\tilde{x}_0 + \tilde{x}_1 + \tilde{x}_2 + ...) = \sum_{t=0}^{\infty} q_{ij,t} = n_{ij},$$

where the last equality comes from (2). *Q.E.D.*

**Definition 2.** Consider an absorbing Markov chain with transition matrix $\Pi$ having the canonical form (1). The matrix $(\mathrm{I} - Q)^{-1}$ is called the **fundamental matrix** for this absorbing chain.

The following theorem states that the sum of the entries in the $i$th row of the fundamental matrix $(\mathrm{I} - Q)^{-1}$ is the expected number of periods before the chain is absorbed when it starts in state $s_i$:

**Theorem 3.** Let $n_i$ be the expected number of periods before the chain is absorbed, given that the chain starts in the transient state $s_i$, and let $n$ be the column vector of dimension $q$ whose $i$th entry is $n_i$. Then, $n = (\mathrm{I} - Q)^{-1}\,\underline{1}$, where $\underline{1}$ is a column vector whose entries are ones.

**Proof.** If we add all the entries in the $i$th row of $(\mathrm{I} - Q)^{-1}$, we will obtain the expected number of times in any of the transient states for a given starting state $s_i$, that is, the expected time required before being absorbed. Thus, $n_i$ is the sum of the entries in the $i$th row of $(\mathrm{I} - Q)^{-1}$. If we write this statement in matrix form, we obtain the theorem. *Q.E.D.*

Our last theorem gives the probability that the chain will end up in a given absorbing state.

**Theorem 4.** Let $b_{ij}$ be the probability that an absorbing chain will be absorbed in the absorbing state $s_j$ if it starts in the transient state $s_i$. Let $B$ be the $q \times d$

matrix with entries $b_{ij}$. Then, $B = (\mathrm{I} - Q)^{-1} D$, where $D$ is as in the canonical form (1).

**Proof.** Let $q_{im,t}$ be the $(i, m)$-entry of the matrix $Q^t$ and $d_{mj}$ the $(m, j)$-entry of the matrix $D$, $i = 1, ..., q$, $m = 1, ..., q$, and $j = 1, ..., d$. The probability $b_{ij}$ may be computed as follows: find the probability of being absorbed by state $s_j$ after spending exactly $t$ steps in transient states and being in the transient state $s_m$ in the last step before absorption. This probability is $q_{im,t}d_{mj}$. Then, add over all transient states before absorption and all number of steps,

$$b_{ij} = \sum_{t=0}^{\infty}\sum_{m=1}^{q} q_{im,t}d_{mj} = \sum_{m=1}^{q}\sum_{t=0}^{\infty} q_{im,t}d_{mj} = \sum_{m=1}^{q} n_{im}d_{mj},$$

where the last equality comes from Theorem 2 and the last expression is the $(i, j)$-entry of the $q \times d$ matrix $(\mathrm{I} - Q)^{-1} D$. *Q.E.D.*

An implication of the previous theorem is that

$$\lim_{t\to\infty} \Pi^t = \widehat{\Pi} = \begin{pmatrix} \widehat{0}_{q\times q} & B_{q\times d} \\ \widehat{0}_{d\times q} & \mathrm{I}_{d\times d} \end{pmatrix},$$

which means that the limiting matrix $\widehat{\Pi}$ exists for absorbing chains.

**Example 1.** Consider the Markov chain with the following transition matrix (Case II in the class notes):

$$\Pi = \begin{pmatrix} Q_{1\times 1} & D_{1\times 2} \\ \widehat{0}_{2\times 1} & \mathrm{I}_{2\times 2} \end{pmatrix} = \begin{pmatrix} 2/5 & 1/5 & 2/5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since the matrix $\Pi$ has the canonical form (1), the Markov chain is absorbing. It has one transient state ($s_1$) and two absorbing states ($s_2$ and $s_2$). Moreover, for this matrix we found that

$$\lim_{t\to\infty} \Pi^t = \widehat{\Pi} = \begin{pmatrix} \widehat{0}_{1\times 1} & B_{1\times 2} \\ \widehat{0}_{2\times 1} & \mathrm{I}_{1\times 2} \end{pmatrix} = \begin{pmatrix} 0 & 1/3 & 2/3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This chain is absorbed with probability 1 (Theorem 1) and the expected number of times the chain will be in the unique transient state $s_1$, given that it starts in $s_1$, is the first (and unique) entry of the $1 \times 1$ fundamental matrix $(\mathrm{I} - Q)^{-1}$,

$$n_1 = \left(1 - \frac{2}{5}\right)^{-1} = \frac{5}{3} = 1.666\,7,$$

as follows from Theorem 2. From Theorem 3 we can compute the expected number of periods before the chain is absorbed, given that the chain starts in the unique transient state $s_1$,

$$(\mathrm{I} - Q)^{-1} \underline{1} = \left(1 - \frac{2}{5}\right)^{-1} \cdot 1 = \frac{5}{3} = 1.6667,$$

which in this case coincides with $n_1$ since there is only one transient state.

Finally, using Theorem 4 we can compute the probabilities that the chain will be absorbed in the absorbing states $s_2$ and $s_3$ if it starts in the transient state $s_1$,

$$B = (\mathrm{I} - Q)^{-1} D = \left(1 - \frac{2}{5}\right)^{-1} \begin{pmatrix} 1/5 & , & 2/5 \end{pmatrix} = \begin{pmatrix} 1/3 & , & 2/3 \end{pmatrix}$$

so that the probabilities that the chain will be absorbed in the absorbing states $s_2$ and $s_3$ if it starts in the transient state $s_1$ are $1/3$ and $2/3$, respectively. This result agrees with the values of the entries of the $1 \times 2$ matrix $B$ appearing in the limiting matrix $\widehat{\Pi}$.

## IRREDUCIBLE MARKOV CHAINS

**Definition 3.** A Markov chain is called **irreducible** (also called **ergodic**) if it is possible to go from every state to every state (not necessarily in one step).

The most obvious questions that can be asked about an irreducible Markov chain are: If the chain is started in state $s_i$, how long will it take on average to return to $s_i$ for the first time? On the average, how long will it take for the chain to reach state $s_j$ for the first time if it started in state $s_i$? The next Theorems 5 and 6 provide the answers to those two questions.

Obviously, a Markov chain is irreducible if its transition matrix $\Pi$ is irreducible, that is, if $\Pi$ cannot be written in the following block upper-triangular form by just renumbering the states:

$$\begin{pmatrix} A & C \\ \widehat{0} & E \end{pmatrix},$$

where $A$ and $E$ are square matrices (possibly of different size so that the matrix $\widehat{0}$ of zeroes does not need to be square).

Remember that a stationary probabilistic row vector satisfies $p^* = p^*\Pi$, that is, $p^*$ is a left-eigenvector associated with the eigenvalue 1. An irreducible chain has a unique stationary probabilistic row vector $p^*$, which is equivalent to say that the irreducible transition matrix $\Pi$ has only one eigenvalue equal to 1 (see the class notes). Moreover, if the chain is irreducible then $p^* \gg 0$.

**Definition 4.** A Markov chain is called **regular** if its transition matrix $\Pi$ is regular (i.e., there exists a natural number $t$ for which the matrix $\Pi^t$ has only strictly positive elements).

In other words, when a chain is regular, for some $t$, it is possible to go from any state to any state in exactly $t$ steps. Therefore, regular Markov chains are irreducible. However, not all irreducible chains are regular. For example, a Markov chain with the transition matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

is irreducible but it is not regular ($\Pi^t$ contains always zeroes for all natural $t$). In this example, it is possible to move from any state to any state but, if $t$ is odd, then it is not possible to move from state $s_1$ to state $s_1$ in $t$ steps, and if $t$ is even, then it is not possible to move from state $s_1$ to state $s_2$ in $t$ steps. Thus, the chain is not regular. For regular Markov chains the ergodic probabilistic vector $\widehat{p}$ exists and, obviously, $\widehat{p} = p^* \gg 0$.

Note that, for a given state $s_j$, if we start in any state an irreducible chain will eventually reach state $s_j$. In fact, it will be in state $s_j$ infinitely often.

**Definition 6.** If an irreducible Markov chain is started in state $s_i$, the expected number of steps to reach state $s_j$ for the first time is called the **mean first passage time** from $s_i$ to $s_j$. It is denoted by $m_{ij}$. By convention, $m_{ii} = 0$.

To find the mean first passage time from $s_i$ to $s_j$ for an irreducible chain we can use the following method: form a new Markov chain by making $s_j$ an absorbing state, that is, define $\pi_{jj} = 1$. If we start at any state other than $s_j$, this new process will behave exactly like the original chain up to the first time that state $s_j$ is reached. Since the original chain was an irreducible chain, it was possible to reach $s_j$ from any other state. Thus the new chain is an absorbing chain with a single absorbing state $s_j$ that will eventually be reached from any other initial state $s_i$ with $i \neq j$. Let $(\mathrm{I} - Q)^{-1}$ be the fundamental matrix for the new absorbing chain. The entries of $(\mathrm{I} - Q)^{-1}$ give then expected number of times in each state before absorption. In terms of the original chain, these quantities give the expected number of times in each of the states before reaching state $s_j$ for the first time. The $i$th component of the $q$ dimension column vector $(\mathrm{I} - Q)^{-1}\,\underline{1}$ gives the expected number of steps before absorption in the new chain, starting in state $s_i$. In terms of the old chain, this is the expected number of steps required to reach state $s_j$ for the first time starting at state $s_i$, i.e., it is the mean first passage time from $s_i$ to $s_j$. This procedure for computing the mean passage time $m_{ij}$ involves the construction of a new absorbing chain for each state $s_j$ to be reached. We will present a more direct procedure in Theorem 6.

Assume that we start in state $s_i$ and consider the length of time before we return to $s_i$ for the first time. If the chain is irreducible, it is clear that we must return, since we either stay at $s_i$ the first step or go to some other state $s_j$, and from any other state $s_j$ we will eventually reach $s_i$. Therefore, all the states of an irreducible Markov chain are recurrent.

**Definition 6.** If an irreducible Markov chain is started in state $s_i$, the expected number of steps to return to $s_i$ for the first time is the **mean recurrence time** for $s_i$. It is denoted by $r_i$.

Consider the mean first passage time from $s_i$ to $s_j$ and assume that $i \neq j$. This may also be computed as follows: take the expected number of steps required given the outcome of the first step, multiply by the probability that this outcome occurs, and add. If the first step is to $s_j$, the expected number of steps required is 1; if it is to some other state $s_k$, the expected number of steps required is $m_{kj}$ plus 1 for the step already taken. Thus,

$$m_{ij} = \pi_{ij} + \sum_{k \neq j} \pi_{ik} \left( m_{kj} + 1 \right),$$

or, since $\sum_{k=1}^{N} \pi_{ik} = 1$,

$$m_{ij} = 1 + \sum_{k \neq j} \pi_{ik} m_{kj}. \tag{3}$$

Similarly, starting in $s_i$, it must take at least one step to return. Considering all possible first steps gives us the mean recurrence time for state $s_i$,

$$r_i = \sum_{k=1}^{N} \pi_{ik} \left( m_{ki} + 1 \right) = 1 + \sum_{k=1}^{N} \pi_{ik} m_{ki}. \tag{4}$$

Let us now define two matrices $M$ and $R$. The $(i,j)$-entry $m_{ij}$ of $M$ is the mean first passage time to go from from $s_i$ to $s_j$ if $i \neq j$ and its diagonal entries are 0. The matrix $M$ is called the mean first passage matrix. The matrix $R$ is a matrix with all entries equal to 0 except the diagonal entries, which are $r_{ii} = r_i$, i.e., the mean recurrence times for each state. The matrix $R$ is called the mean recurrence matrix. Let $\widehat{1}$ be an $N \times N$ matrix with all entries equal to 1. Using (3) for the case $i \neq j$ and (4) for the case $i = j$, we obtain the following matrix equation:

$$M + R = \Pi M + \widehat{1}$$

or

$$\left( \mathrm{I} - \Pi \right) M = \widehat{1} - R. \tag{5}$$

The previous equation with $m_{ii} = 0$ implies (3) and (4). The next theorem gives the mean recurrence time for a given state.

**Theorem 5.** For an irreducible Markov chain, the mean recurrence time for state $s_i$ is $r_i = 1/p_i^*$, where $p_i^*$ is the $i$th component of the unique stationary probabilistic row vector $p^*$ for the transition matrix $\Pi$.

**Proof.** Multiplying both sides of (5) by $p^*$, and using the fact that

$$p^*(\mathrm{I} - \Pi) = \underline{0}^\intercal,$$

we get

$$p^*\widehat{1} - p^* R = \underline{0}^\intercal.$$

Here $p^*\widehat{1}$ is a row vector with all entries equal to 1 (since $p^*$ is a probabilistic vector) and $p^* R$ is a row vector with $i$th entry equal $p_i^* r_i$ (since $R$ is a diagonal matrix). Thus,

$$(1, 1, ..., 1) = (p_1^* r_1, p_2^* r_2, ..., p_N^* r_N)$$

so that $r_i = 1/p_i^*$, as was to be proved. *Q.E.D.*

A direct corollary of the previous theorem is that, for an irreducible Markov chain, the entries of the stationary probabilistic row vector $p^*$ are strictly positive. This is so because we know that the values of $r_i$ are finite and, hence, $p_i^* = 1/r_i$ cannot be equal to 0.

We next present a more direct procedure to calculate the mean passage matrix $M$ based on the use of the fundamental matrix for irreducible Markov chains. In order to state the main result in Theorem 6 we need two previous results (a proposition and a lemma). The proofs of these Proposition and Lemma rely only on the assumption that the transition matrix $\Pi$ of the Markov chain has a only one eigenvalue equal to 1 or, equivalently, that the chain has only one stationary probabilistic row vector $p^*$. This assumption is satisfied by all irreducible Markov chains.

**Proposition** (Existence of the fundamental matrix). Assume that the Markov chain has only one stationary probabilistic row vector $p^*$ and let $\Pi^*$ be a stochastic matrix all of whose rows are the stationary probabilistic row vector $p^*$. The matrix $\mathrm{I} - \Pi + \Pi^*$ has an inverse $(\mathrm{I} - \Pi + \Pi^*)^{-1}$.

**Proof.** Let $x$ be a column vector such that

$$(\mathrm{I} - \Pi + \Pi^*)\, x = \underline{0}. \tag{6}$$

To prove the proposition, it is sufficient to show that $x$ must be the zero vector. Multiplying (6) by the stationary probabilistic vector $p^*$ and using the fact that $p^*$ satisfies $p^*(\mathrm{I} - \Pi) = \underline{0}^\intercal$ and $p^*\Pi^* = p^*$, we get

$$p^*\,(\mathrm{I} - \Pi + \Pi^*)\, x = p^* x = 0.$$

Therefore, since the entries of the vector $\Pi^* x$ are all $p^* x = 0$, we have that $\Pi^* x = \underline{0}$. Thus, (6) becomes simply $(\mathrm{I} - \Pi)\, x = \underline{0}$ or $(\Pi - \mathrm{I})\, x = \underline{0}$ so that $x$ is a column vector belonging to the (right-)eigenspace of the matrix $\Pi$. Remember from the class notes that, since the irreducible transition matrix $\Pi$ has only one eigenvalue equal to 1, the (right-)eigenspace associated with this eigenvalue consist of all the column vectors with identical values in its cells $x = (\alpha, \alpha, ..., \alpha)^\intercal$. This means that all the entries of the vector $x$ must have the same sign. Since $p^* > 0$ and $p^* x = 0$, then $x = \underline{0}$.[1] *Q.E.D.*

**Definition 7.** Consider an irreducible Markov chain with the transition matrix $\Pi$ and the stationary probabilistic row vector $p^*$. The matrix $(\mathrm{I} - \Pi + \Pi^*)^{-1}$ is called the **fundamental matrix** for this irreducible Markov chain.

Note that the fundamental matrix for an irreducible Markov chain, $(\mathrm{I} - \Pi + \Pi^*)^{-1}$, has not the same expression as the fundamental matrix for an absorbing Markov chain, $(\mathrm{I} - Q)^{-1}$.

**Lemma** (Properties of the fundamental matrix).

$$(\mathrm{I} - \Pi + \Pi^*)^{-1}\, \underline{1} = \underline{1} \ \text{ or, equivalently, } \ (\mathrm{I} - \Pi + \Pi^*)^{-1}\, \widehat{1} = \widehat{1} \qquad \text{(a)}$$

$$p^* \left(\mathrm{I} - \Pi + \Pi^*\right)^{-1} = p^*, \quad \text{or, equivalently, } \ \Pi^* \left(\mathrm{I} - \Pi + \Pi^*\right)^{-1} = \Pi^* \qquad \text{(b)}$$

and

$$(\mathrm{I} - \Pi + \Pi^*)^{-1} (\mathrm{I} - \Pi) = \mathrm{I} - \Pi^*. \qquad \text{(c)}$$

**Proof.** Since, $\mathrm{I}\underline{1} = \underline{1}$, $\Pi\underline{1} = \underline{1}$ and $\Pi^*\underline{1} = \underline{1}$,

$$\underline{1} = (\mathrm{I} - \Pi + \Pi^*)\, \underline{1}.$$

If we premultiply both sides of this equation by $(\mathrm{I} - \Pi + \Pi^*)^{-1}$, we obtain (a).

Similarly, since $p^*\mathrm{I} = p^*$, $p^*\Pi = p^*$ and $p^*\Pi^* = p^*$,

$$p^* = p^* \left(\mathrm{I} - \Pi + \Pi^*\right).$$

If we postmultiply both sides of this equation by $(\mathrm{I} - \Pi + \Pi^*)^{-1}$, we obtain (b).

Finally, since both $\Pi$ and $\Pi^*$ are stochastic matrices and $\Pi^*$ has identical rows, we have that $\Pi\Pi^* = \Pi^*$ and $\Pi^*\Pi^* = \Pi^*$. Therefore,

$$\begin{aligned}(\mathrm{I} - \Pi + \Pi^*)(\mathrm{I} - \Pi^*) &= \mathrm{I} - \Pi^* - \Pi + \Pi\Pi^* + \Pi^* - \Pi^*\Pi^* \\ &= \mathrm{I} - \Pi^* - \Pi + \Pi^* + \Pi^* - \Pi^* = \mathrm{I} - \Pi.\end{aligned}$$

If we premultiply both sides of this equation by $(\mathrm{I} - \Pi + \Pi^*)^{-1}$, we obtain (c). *Q.E.D.*

[1] Remember that $p^* > 0$ means that $p_i \geq 0$ for $i = 1, ..., N$ and $p_i > 0$ for at least one $i$. This has to be the case for a probabilistic vector since $\sum_{i=1}^{N} p_i = 1$.

**Theorem 6.** The mean first passage matrix $M$ for an irreducible chain is determined from the fundamental matrix $(\mathrm{I}-\Pi+\Pi^*)^{-1}$ and the stationary probabilistic row vector $p^*$ by

$$m_{ij} = \frac{z_{jj} - z_{ij}}{p_j^*}, \tag{7}$$

where $z_{ij}$ is the $(i,j)$-entry of the fundamental matrix $(\mathrm{I}-\Pi+\Pi^*)^{-1}$ and $p_j^*$ is the $j$th entry of the stationary probabilistic row vector $p^*$.

**Proof.** Premultiplying both sides of (5) by the fundamental matrix $(\mathrm{I}-\Pi+\Pi^*)^{-1}$ we get

$$(\mathrm{I}-\Pi+\Pi^*)^{-1}(\mathrm{I}-\Pi)\,M = (\mathrm{I}-\Pi+\Pi^*)^{-1}\,\widehat{1} - (\mathrm{I}-\Pi+\Pi^*)^{-1}\,R,$$

and from (a) and (c) in the Lemma, the previous equation becomes

$$(\mathrm{I}-\Pi^*)\,M = \widehat{1} - (\mathrm{I}-\Pi+\Pi^*)^{-1}\,R$$

or

$$M = \widehat{1} - (\mathrm{I}-\Pi+\Pi^*)^{-1}\,R + \Pi^* M.$$

From this equation and the definition of the diagonal matrix $R$, we see that

$$m_{ij} = 1 - z_{ij} r_j + (p^* M)_j\,, \tag{8}$$

where $(p^* M)_j$ is the $j$th entry of the row vector $p^* M$. Since $m_{jj} = 0$, the previous expression becomes

$$0 = 1 - z_{jj} r_j + (p^* M)_j$$

or

$$(p^* M)_j = z_{jj} r_j - 1. \tag{9}$$

Plugging (9) into (8), we have

$$m_{ij} = (z_{jj} - z_{ij})\, r_j.$$

and, since from Theorem 5 $r_i = 1/p_i^*$, we obtain (7). *Q.E.D.*

**Example 2.** Consider the Markov chain with the following regular transition matrix (Case I in the class notes):

$$\Pi = \begin{pmatrix} 3/10 & 0 & 7/10 \\ 4/5 & 1/5 & 0 \\ 0 & 3/5 & 2/5 \end{pmatrix}.$$

For this example we found that

$$p^* = \widehat{p} = \left(\frac{24}{73}, \frac{21}{73}, \frac{28}{73}\right) \gg 0.$$

Therefore, for instance, the mean recurrence time for $s_2$ is $r_2 = 1/p_2^* = 73/21 = 3.4762$. We can also compute the fundamental matrix for this chain,

$$(\mathrm{I} - \Pi + \Pi^*)^{-1} = \begin{pmatrix} 1 - \frac{3}{10} + \frac{24}{73} & 0 - 0 + \frac{21}{73} & 0 - \frac{7}{10} + \frac{28}{73} \\ 0 - \frac{4}{5} + \frac{24}{73} & 1 - \frac{1}{5} + \frac{21}{73} & 0 - 0 + \frac{28}{73} \\ 0 - 0 + \frac{24}{73} & 0 - \frac{3}{5} + \frac{21}{73} & 1 - \frac{2}{5} + \frac{28}{73} \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \frac{751}{730} & \frac{21}{73} & -\frac{231}{730} \\ -\frac{172}{365} & \frac{397}{365} & \frac{28}{73} \\ \frac{24}{73} & -\frac{114}{365} & \frac{359}{365} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{4342}{5329} & -\frac{672}{5329} & \frac{1659}{5329} \\ \frac{2152}{5329} & \frac{4073}{5329} & -\frac{896}{5329} \\ -\frac{768}{5329} & \frac{1518}{5329} & \frac{4579}{5329} \end{pmatrix}.$$

So, for example, the mean first passage time from $s_1$ to $s_2$, is

$$m_{12} = \frac{z_{22} - z_{12}}{p_2^*} = \frac{\frac{4073}{5329} - \left(-\frac{672}{5329}\right)}{\frac{21}{73}} = \frac{65}{21} = 3.0952.$$

**Example 3.** Consider the Markov chain with the following transition matrix (Case III in the class notes):

$$\Pi = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 3/10 & 2/10 & 1/2 \end{pmatrix}.$$

This chain is not irreducible since, if the chain starts in either state $s_1$ or state $s_2$, it will never reach state $s_3$. The set $\{s_1, s_2\}$ is ergodic, whereas the state $s_3$ is transient. Note that the previous matrix it is not irreducible since by permuting states $s_1$ and $s_3$ we obtain the following block upper-triangular matrix:

$$\begin{pmatrix} A & C \\ \widehat{0} & E \end{pmatrix} = \begin{pmatrix} 1/2 & 2/10 & 3/10 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

For this example we found that the matrix $\Pi$ has the eigenvalues $1, -1$ and $0.5$ so that it has a unique eigenvalue equal to 1. Therefore, the chain has a unique stationary probabilistic row vector, which is

$$p^* = \left(\frac{1}{2}, \frac{1}{2}, 0\right).$$

However, the stationary probabilistic vector does not have all its entries strictly positive.

Finally, this Markov chain is not absorbing either: it has no absorbing states and its transition matrix $\Pi$ cannot be rewritten as in the canonical form (1) since the matrix $E$ is not the identity matrix I. Nevertheless, the chain ends up being absorbed by the ergodic set $\{s_1, s_2\}$ and, after being absorbed, it oscillates between states $s_1$ and $s_2$ forever.

The Poisson distribution as a limit of binomial distributions

$$\lim_{\substack{n\to\infty \\ n\theta=\lambda\in(0,\infty)}} b(x;n,\theta) = \lim_{\substack{\theta\to 0 \\ n\theta=\lambda\in(0,\infty)}} b(x;n,\theta) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \text{for } x = 0,\ 1,\ \ldots$$

**Proof:** Since $\theta = \lambda/n$,

$$b(x;n,\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x(1-\theta)^{n-x}$$

$$= \frac{n(n-1)(n-2)\cdot\cdots\cdot(n-x+1)}{x!}\frac{\lambda^x}{n^x}\left[(1-\theta)^{-n/\lambda}\right]^{-\lambda}(1-\theta)^{-x}$$

$$= 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdot\cdots\cdot\left(1-\frac{x-1}{n}\right)\frac{\lambda^x}{x!}\left[(1-\theta)^{-1/\theta}\right]^{-\lambda}(1-\theta)^{-x}.$$

Therefore,

$$\lim_{\substack{n\to\infty \\ \theta\to 0 \\ n\theta=\lambda\in(0,\infty)}} b(x;n,\theta) = 1\cdot\frac{\lambda^x}{x!}\cdot\lim_{\theta\to 0}\left[(1-\theta)^{-1/\theta}\right]^{-\lambda}\cdot 1$$

$$= \frac{\lambda^x}{x!}\cdot\lim_{\theta\to 0}\left[(1-\theta)^{-1/\theta}\right]^{-\lambda} = \frac{\lambda^x}{x!}\left[\lim_{\theta\to 0}(1-\theta)^{-1/\theta}\right]^{-\lambda} = \frac{\lambda^x}{x!}e^{-\lambda},$$

where the last equality follows from

$$\lim_{\theta\to 0}(1-\theta)^{-1/\theta} = \lim_{\theta\to 0}\left(e^{\ln\left[(1-\theta)^{-1/\theta}\right]}\right) = \lim_{\theta\to 0}\left[e^{-\frac{1}{\theta}\ln(1-\theta)}\right] = e^{-\lim_{\theta\to 0}\left[\frac{1}{\theta}\ln(1-\theta)\right]} = e$$

since, using the l'Hôpital's rule, we can show that $\lim_{\theta\to 0}\left[\frac{1}{\theta}\ln(1-\theta)\right] = -1$. *Q.E.D.*

− − − − − − − − − − − − − − − − − − −

The random variable $\tilde{x}$ has a Poisson distribution if its probability function is

$$p(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad \text{with } \lambda > 0, \quad \text{for } x = 0,\ 1,\ \ldots$$

We write $\widetilde{x} \sim \mathrm{P}(\lambda)$.

Thus, if $\widetilde{x}_n \sim \mathrm{B}(n,\theta)$, then $x_n \to \mathrm{P}(\lambda)$ when $n \to \infty$ and $n\theta = \lambda \in (0,\infty)$ (or when $\theta \to 0$ and $n\theta = \lambda \in (0,\infty)$).

– – – – – – – – – – – – – – – – – –

Mean and variance of the Poisson distribution:

$$\mu = n\theta = \lambda,$$

and

$$\sigma^2 = n\theta(1-\theta) = \lambda,$$

since $n\theta = \lambda$ and $\theta \to 0$.

– – – – – – – – – – – – – – – – – –

**Moment-generating function of the Poisson distribution,** $M_{\tilde{x}}(t)$.

We compute the MGF $M_{\tilde{x}}(t)$ as the limit of the moment-generating function of the binomial distribution, $[1+\theta(e^t-1)]^n$.

$$M_{\tilde{x}}(t) = \lim_{\substack{n\to\infty \\ n\theta=\lambda\in(0,\infty)}} \left[1+\theta(e^t-1)\right]^n = \lim_{n\to\infty}\left[1+\frac{\lambda}{n}(e^t-1)\right]^n$$

$$= \lim_{n\to\infty}\left(\left[1+\frac{1}{\frac{n}{\lambda(e^t-1)}}\right]^{\frac{n}{\lambda(e^t-1)}}\right)^{\lambda(e^t-1)} = \lim_{z\to\infty}\left[\left(1+\frac{1}{z}\right)^z\right]^{\lambda(e^t-1)}$$

$$= \left[\lim_{z\to\infty}\left(1+\frac{1}{z}\right)^z\right]^{\lambda(e^t-1)} = e^{\lambda(e^t-1)},$$

where $z = \dfrac{n}{\lambda(e^t-1)} \to \infty$ as $n \to \infty$.

– – – – – – – – – – – – – – – – – –

*Exercise:* Check that $M_{\tilde{x}}(t) = \sum_{x=0}^{\infty} e^{tx} p(x;\,\lambda) = e^{\lambda(e^t-1)}$ and use $M_{\tilde{x}}(t)$ to find $\mu$ and $\sigma^2$.

– – – – – – – – – – – – – – – – – –

## The standard normal distribution as a limit of standardized binomial distributions

**Theorem (De Moivre-Laplace).** If $\tilde{x}_n$ is a random variable having a binomial distribution with parameters $n$ and $\theta \in (0,1)$, i.e., its probability function is $b(x; n, \theta)$, then the moment generating function of the standardization of $\tilde{x}_n$,

$$\tilde{z}_n = \frac{\widetilde{x}_n - \mathrm{E}\left(\widetilde{x}_n\right)}{\sqrt{\mathrm{Var}\left(\widetilde{x}_n\right)}} = \frac{\tilde{x}_n - n\theta}{\sqrt{n\theta(1-\theta)}}$$

tends to that of the standard normal distribution as $n \to \infty$. Hence, $\tilde{z}_n \to \mathrm{N}(0,1)$.

**Proof:** Let $n\theta \equiv \mu$ and $n\theta(1-\theta) \equiv \sigma^2$. Then we know that

$$M_{\tilde{z}_n}(t) = M_{\frac{\tilde{x}_n - \mu}{\sigma}}(t) = e^{-\frac{\mu t}{\sigma}} M_{\tilde{x}_n}\left(\frac{t}{\sigma}\right) = e^{-\frac{\mu t}{\sigma}}[1 + \theta(e^{t/\sigma} - 1)]^n.$$

Then, taking logarithms and substituting the MacLaurin series (or Taylor expansion around zero) of $e^{t/\sigma}$, we get

$$\ln M_{\tilde{z}_n}(t) = -\frac{\mu t}{\sigma} + n \ln[1 + \theta(e^{t/\sigma} - 1)]$$

$$= -\frac{\mu t}{\sigma} + n \ln\left[1 + \underbrace{\theta\left\{\overbrace{\frac{1}{1!}\left(\frac{t}{\sigma}\right) + \frac{1}{2!}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{3!}\left(\frac{t}{\sigma}\right)^3 + \ldots}^{e^{t/\sigma}-1}\right\}}_{x}\right]$$

and using the infinite series $\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \ldots$, which converges for $|x| < 1$, to expand this logarithm, it follows that

$$\ln M_{\tilde{z}_n}(t) = -\frac{\mu t}{\sigma} + n\theta\left[\frac{t}{\sigma} + \frac{1}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{6}\left(\frac{t}{\sigma}\right)^3 + \ldots\right]$$

$$-\frac{n\theta^2}{2}\left[\frac{t}{\sigma} + \frac{1}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{6}\left(\frac{t}{\sigma}\right)^3 + \ldots\right]^2$$

$$+\frac{n\theta^3}{3}\left[\frac{t}{\sigma} + \frac{1}{2}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{6}\left(\frac{t}{\sigma}\right)^3 + \ldots\right]^3 - \ldots$$

Collecting powers of $t$, we obtain

$$\ln M_{\tilde{z}_n}(t) = \underbrace{\left(-\frac{\mu}{\sigma} + \frac{n\theta}{\sigma}\right)}_{=0} t + \left(\frac{n\theta}{2\sigma^2} - \frac{n\theta^2}{2\sigma^2}\right) t^2$$

$$+ \left(\frac{n\theta}{6\sigma^3} - \frac{n\theta^2}{2\sigma^3} + \frac{n\theta^3}{3\sigma^3}\right) t^3 + \dots$$

$$= \frac{1}{\sigma^2}\left(\frac{n\theta - n\theta^2}{2}\right) t^2 + \frac{n}{\sigma^3}\left(\frac{\theta - 3\theta^2 + 2\theta^3}{6}\right) t^3 + \dots$$

since $\mu = n\theta$. Then substituting $\sigma^2 = n\theta(1-\theta) = n\theta - n\theta^2$, we find that

$$\ln M_{\tilde{z}_n}(t) = \frac{1}{2}t^2 + \frac{n}{\sigma^3}\left(\frac{\theta - 3\theta^2 + 2\theta^3}{6}\right) t^3 + \dots,$$

where for $r > 2$ the coefficient of $t^r$ is a constant times $\frac{n}{\sigma^r}$, which approaches 0 when $n \to \infty$ since

$$\frac{n}{\sigma^r} = \frac{n}{n^{r/2}[\theta(1-\theta)]^{r/2}}.$$

It follows that

$$\lim_{n\to\infty} [\ln M_{\tilde{z}_n}(t)] = \frac{1}{2}t^2$$

and, since the limit of the logarithm equals the logarithm of the limit (provided that the limit exists), we conclude that

$$\lim_{n\to\infty} M_{\tilde{z}_n}(t) = e^{\frac{1}{2}t^2},$$

which is the moment-generating function of a standard normal random variable. *Q.E.D.*

# Central Limit Theorem (Lindeberg-Levy)

**Theorem.** Let $\{\widetilde{x}_i\}_{i=1}^{\infty}$ be i.i.d. random variables with $\mathrm{E}(\widetilde{x}_i) = \mu$ and $\mathrm{Var}(\widetilde{x}_i) = \sigma^2$, with $0 < \sigma^2 < \infty$ , for all $i$. Then

$$\widetilde{z}_n \equiv \frac{\overline{\boldsymbol{x}}_n - \mathrm{E}(\overline{\boldsymbol{x}}_n)}{\sqrt{\mathrm{Var}(\overline{\boldsymbol{x}}_n)}} = \frac{\overline{\boldsymbol{x}}_n - \mu}{\sigma / \sqrt{n}}$$

converges in distribution to a standard normal random variable as $n \to \infty$. That is, $\widetilde{z}_n \longrightarrow \mathrm{N}(0, 1)$.

**Proof.** We will assume in this proof that the common distribution of all the random variables in the sequence $\{\widetilde{x}_i\}_{i=1}^{\infty}$ has a moment-generating function $M_{\widetilde{x}}(t)$ that is well-defined (i.e., finite) in a neighborhood of $t = 0$. Then,

$$M_{\widetilde{z}_n}(t) = M_{\frac{\overline{\boldsymbol{x}}_n - \mu}{\sigma/\sqrt{n}}}(t) = e^{-\frac{\sqrt{n}\cdot\mu t}{\sigma}} M_{\overline{\boldsymbol{x}}_n}\left(\frac{\sqrt{n}\cdot t}{\sigma}\right) = e^{-\frac{\sqrt{n}\cdot\mu t}{\sigma}} M_{n\cdot\overline{\boldsymbol{x}}_n}\left(\frac{t}{\sigma\sqrt{n}}\right).$$

Since $n \cdot \overline{\boldsymbol{x}}_n = \widetilde{x}_1 + \widetilde{x}_2 + \ldots + \widetilde{x}_n$ and $\{\widetilde{x}_i\}_{i=1}^{\infty}$ are i.i.d., it follows that

$$M_{\widetilde{z}_n}(t) = e^{-\frac{\sqrt{n}\cdot\mu t}{\sigma}} \left[M_{\widetilde{x}}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n,$$

where $M_{\widetilde{x}}(t) \equiv M_{\widetilde{x}_i}(t)$, for all $i$. Hence,

$$\ln M_{\widetilde{z}_n}(t) = -\frac{\sqrt{n}\cdot\mu t}{\sigma} + n \ln M_{\widetilde{x}}\left(\frac{t}{\sigma\sqrt{n}}\right).$$

Expanding $M_{\widetilde{x}}\left(\frac{t}{\sigma\sqrt{n}}\right)$ as a power series in $t$, we obtain

$$\ln M_{\widetilde{z}_n}(t) = -\frac{\sqrt{n}\cdot\mu t}{\sigma} + n \ln\left[1 + \mu_1' \frac{t}{\sigma\sqrt{n}} + \mu_2' \frac{t^2}{2!\cdot\sigma^2 n} + \mu_3' \frac{t^3}{3!\cdot\sigma^3 n^{3/2}} + \ldots\right],$$

where $\mu_1'$, $\mu_2'$, $\mu_3', \ldots$ are the (non-central) moments of $\widetilde{x}_i$, for all $i$.

If $n$ is sufficiently large, we can use the expansion of $\ln(1+x)$ as a power series in $x$, $\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \ldots$, getting

$$\begin{aligned}\ln M_{\widetilde{z}_n}(t) = -\frac{\sqrt{n}\cdot\mu t}{\sigma} + n\Bigg\{ &\left[\mu_1' \frac{t}{\sigma\sqrt{n}} + \mu_2' \frac{t^2}{2\sigma^2 n} + \mu_3' \frac{t^3}{6\sigma^3 n^{3/2}} + \ldots\right] \\ &-\frac{1}{2}\left[\mu_1' \frac{t}{\sigma\sqrt{n}} + \mu_2' \frac{t^2}{2\sigma^2 n} + \mu_3' \frac{t^3}{6\sigma^3 n^{3/2}} + \ldots\right]^2 \\ &+\frac{1}{3}\left[\mu_1' \frac{t}{\sigma\sqrt{n}} + \mu_2' \frac{t^2}{2\sigma^2 n} + \mu_3' \frac{t^3}{6\sigma^3 n^{3/2}} + \ldots\right]^3 - \ldots\Bigg\}\end{aligned}$$

Then collecting powers of $t$, we obtain

$$\ln M_{\tilde{z}_n}(t) = \left(-\frac{\sqrt{n}\cdot\mu}{\sigma} + \frac{\sqrt{n}\cdot\mu_1'}{\sigma}\right)t + \left(\frac{\mu_2'}{2\sigma^2} - \frac{(\mu_1')^2}{2\sigma^2}\right)t^2$$

$$+\left(\frac{\mu_3'}{6\sigma^3\sqrt{n}} - \frac{\mu_1'\mu_2'}{2\sigma^3\sqrt{n}} + \frac{(\mu_1')^3}{3\sigma^3\sqrt{n}}\right)t^3 + \ldots,$$

and, since $\mu_1' = \mu$ and $\mu_2' - (\mu_1')^2 = \sigma^2$, this reduces to

$$\ln M_{\tilde{z}_n}(t) = \frac{1}{2}t^2 + \left(\frac{\mu_3'}{6} - \frac{\mu_1'\mu_2'}{2} + \frac{\mu_1'^3}{3}\right)\frac{t^3}{\sigma^3\sqrt{n}} + \ldots$$

Finally, observing that the coefficient of $t^3$ is a constant times $\frac{1}{\sqrt{n}}$ and, in general, the coefficient of $t^r$ is a constant times $\frac{1}{n^{(r-2)/2}}$, we get

$$\lim_{n\to\infty}\left[\ln M_{\tilde{z}_n}(t)\right] = \frac{1}{2}t^2$$

and, hence,

$$\lim_{n\to\infty} M_{\tilde{z}_n}(t) = e^{\frac{1}{2}t^2},$$

which is the moment-generating function of a standard normal random variable $\tilde{z}$. Therefore, since $\lim_{n\to\infty} M_{\tilde{z}_n}(t) = M_{\tilde{z}}(t)$, it follows that

$$\tilde{z}_n \xrightarrow{d} \tilde{z} \quad \text{or} \quad \tilde{z}_n \to \mathrm{N}(0,1). \qquad \textit{Q.E.D.}$$

**Note:** An almost identical proof could be made using the characteristic function instead of the moment-generating function. Notice that the characteristic function is finite everywhere for all random variables.

**Exercises. Probability and Statistics. IDEA.**
**6. Stochastic Processes and Limiting Distributions**

1. Consider the stochastic process $\{\tilde{x}_t\}_{t=0}^{T}$ of random variables on the probability space $(\Omega, \mathcal{F}, P)$ adapted to the filtration $\mathbb{F}$. Prove the following two theorems:

   $(a)$ Let $\{\tilde{x}_t\}_{t=0}^{T}$ be a submartingale and $g$ a convex, increasing function from $\mathbb{R}$ to $\mathbb{R}$. If $g(\tilde{x}_t)$ is integrable for all $t$, then $\{g(\tilde{x}_t)\}_{t=0}^{T}$ is a submartingale.

   $(b)$ Let $\{\tilde{x}_t\}_{t=0}^{T}$ be a martingale and $g$ a convex function from $\mathbb{R}$ to $\mathbb{R}$. If $g(\tilde{x}_t)$ is integrable for all $t$, then $\{g(\tilde{x}_t)\}_{t=0}^{T}$ is a submartingale. Thus, if $r \geq 1$, $\{\tilde{x}_t\}_{t=0}^{T}$ is a martingale and $|\tilde{x}_t|^r$ is integrable for all $t$, then $\{|\tilde{x}_t|^r\}_{t=0}^{T}$ is a submartingale.

2. Imagine that you flip a balanced coin 16 times.

   $(a)$ Find the probability of obtaining exactly 6 heads. Solve this problem using both the binomial distribution and the corresponding approximation by the standard normal. Compare the results.

   $(b)$ Find the probability of obtaining a number of heads strictly higher than 4 and smaller or equal than 7. Solve this problem using both the binomial distribution and the corresponding approximation by the standard normal. Compare the results.

3. Find the probability that exactly 10 individuals pass a (very demanding) exam if 3000 individuals have taken the exam and the probability of passing is just 0.005. Solve this problem using both the binomial distribution and the corresponding approximation by the Poisson distribution. Compare the results.

4. Consider a Markov chain $\{\tilde{x}_t\}_{t=0}^{\infty}$ on the probability space $(\Omega, \mathcal{F}, P)$. The state space is $S = \{s_1, s_2, s_3\}$ and the time-invariant transition matrix is

$$\begin{pmatrix} 1 & 0 & 0 \\ 1/6 & 1/6 & 2/3 \\ 0 & 0 & 1 \end{pmatrix},$$

   where the value appearing in the $(i,j)$-entry of the transition matrix is the conditional probability $\pi_{ij} = P\{\tilde{x}_{t+1} = s_j | \tilde{x}_t = s_i\}$ for all $t$.

   $(a)$ Find the stationary probabilistic vector (or vectors, if there are more than one), the ergodic probabilistic vector (if it exists), the transient states (if they exist), the recurrent states (if they exist), the absorbing states (if they exist) and the ergodic sets of the Markov chain.

   $(b)$ Assume that the Markov chain starts in state $s_2$, $\tilde{x}_0 = s_2$. Compute

$$\lim_{t\to\infty} P\{\tilde{x}_t = s_1 | \tilde{x}_0 = s_2\}, \ \lim_{t\to\infty} P\{\tilde{x}_t = s_2 | \tilde{x}_0 = s_2\}, \text{ and } \lim_{t\to\infty} P\{\tilde{x}_t = s_3 | \tilde{x}_0 = s_2\}.$$

   $(c)$ We flip simultaneously three balanced coins. If we obtain an odd number (one or three) of heads, the initial state is $s_1$ ($\tilde{x}_0 = s_1$); if we get two heads, the initial state is $s_2$ ($\tilde{x}_0 = s_2$); and if we get no heads, the initial state is $s_3$ ($\tilde{x}_0 = s_3$). What is the unconditional (i.e., before flipping the coins) distribution of states in the long run?

($d$) Find the expected number of periods (or steps) before the Markov chain is absorbed if it has started in state $s_2$?

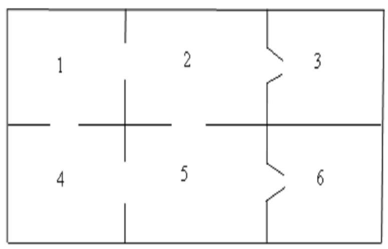
5. Consider the following time-invariant transition matrices:

$$(a)\, \mathrm{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (b) \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad (c) \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad (d) \begin{pmatrix} 1/2 & 1/2 \\ 3/4 & 1/4 \end{pmatrix}$$

($i$) Find the stationary probabilistic vectors, the ergodic probabilistic vectors (if they exist), the transient states (if they exist), the recurrent states (if they exist), the absorbing states (if they exist) and the ergodic sets of the Markov chains having the previous transition matrices.

($ii$) Consider the Markov chain with the transition matrix ($d$). Find the mean recurrence time for each of the two states of the Markov chain and the mean first passage time from state $s_1$ to $s_2$ and from state $s_2$ to $s_1$.

6. A mouse is in the following labyrinth:




In each period the mouse moves randomly from one cell to an adjacent one. All the adjacent cells are equally likely. However, if the mouse enters into cells 3 or 6, it cannot exit from them.

($a$) Find the probability that the mouse will enter into cell 6 if it was initially in cell 1.

($b$) Find the expected number of periods the mouse will be in cell 2 given that it was initially in cell 5.

($c$) Find the expected number of periods before the mouse is trapped if it was initially in cell 4.

7. Two players $a$ and $b$ have two dollars each at the beginning of the game. They bet on repeated flips of a coin. At each flip, the loser pays the winner one dollar, and the game continues until either player is "ruined".

($a$) Find the expected duration of the game

($b$) Find the probability that $a$ will win the game after flipping the coin exactly four times.

($c$) If we know that player $a$ has three dollars in a given moment of the game, what is the probability that player $b$ ends up winning the game?

8. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Assume that the sample space $\Omega$ is the closed interval of real numbers $[0,1]$, $\mathcal{F}$ is the $\sigma-$algebra of Borel sets in the interval $[0,1]$, and the probability $P$ on $(\Omega, \mathcal{F})$ is the Lebesgue measure on $[0,1]$. Consider the sequence of random variables $\{\tilde{x}_n\}_{n=1}^{\infty}$ on $(\Omega, \mathcal{F}, P)$ defined as follows:

$$\tilde{x}_1(\omega) = \begin{cases} 1 & \text{if } \omega \in [0,1] \\ 0 & \text{otherwise,} \end{cases}$$

$$\tilde{x}_2(\omega) = \begin{cases} 1 & \text{if } \omega \in \left[0, \frac{1}{2}\right] \\ 0 & \text{otherwise,} \end{cases}$$

$$\tilde{x}_3(\omega) = \begin{cases} 1 & \text{if } \omega \in \left[\frac{1}{2}, \frac{1}{2} + \frac{1}{3}\right] \\ 0 & \text{otherwise,} \end{cases}$$

$$\tilde{x}_4(\omega) = \begin{cases} 1 & \text{if } \omega \in \left[0, \frac{1}{12}\right] \cup \left[\frac{1}{2} + \frac{1}{3}, 1\right] \\ 0 & \text{otherwise,} \end{cases}$$

$$\tilde{x}_5(\omega) = \begin{cases} 1 & \text{if } \omega \in \left[\frac{1}{12}, \frac{1}{12} + \frac{1}{5}\right] \\ 0 & \text{otherwise,} \end{cases}$$

and so on. In other words, the subset of $[0,1]$ over which $\tilde{x}_n$ is equal to 1 has the length $\frac{1}{n}$ and keeps moving to the right until it reaches the right end point of $[0,1]$, at which point it moves back to 0 and starts again.

Does $\tilde{x}_n$ converges almost surely to the constant 0 as $n \to \infty$? Does $\tilde{x}_n$ converges in probability to the constant 0 as $n \to \infty$? Does $\tilde{x}_n$ converges in distribution to the constant 0 as $n \to \infty$?

*Hint:* Remember that $\sum_{i=1}^{\infty} \frac{1}{i} = \infty$.

9. Consider a Markov chain $\{\tilde{x}_t\}_{t=0}^{\infty}$ on the probability space $(\Omega, \mathcal{F}, P)$. The state space is $S = \{s_1, s_2, s_3\}$ and the time-invariant transition matrix is

$$\Pi = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 3/10 & 0 & 7/10 \end{pmatrix},$$

where the value appearing in the $(i, j)$-entry of the matrix $\Pi$ is the conditional probability $\pi_{ij} = P\{\tilde{x}_{t+1} = s_j | \tilde{x}_t = s_i\}$ . for all $t$.

$(a)$ Find the stationary probabilistic vector (or vectors, if there are more than one), the ergodic probabilistic vector (if it exists), the transient states (if they exist), the recurrent states (if they exist), the absorbing states (if they exist) and the ergodic sets of the Markov chain.

$(b)$ Compute the matrix $\widehat{\Pi} = \lim_{t\to\infty} \Pi^t$. *Hint:* Recall that $\Pi \cdot \widehat{\Pi} = \lim_{t\to\infty} \Pi^{t+1} = \widehat{\Pi}$ and, equivalently, $\widehat{\Pi} \cdot \Pi = \lim_{t\to\infty} \Pi^{t+1} = \widehat{\Pi}$.

$(c)$ We roll a dice and, if we get either 1 or 5 dots, the Markov chain starts in state $s_1$; if we get an even number of dots, it starts in state $s_2$; and if we get 3 dots, it starts in state $s_3$. What is the unconditional (i.e., before rolling the dice) distribution of states in the long run?

10. Consider a Markov chain $\{\tilde{x}_t\}_{t=0}^{\infty}$ on the probability space $(\Omega, \mathcal{F}, P)$. The state space is $S = \{s_1, s_2, s_3, s_4\}$ and the time-invariant transition matrix is

$$\Pi = \begin{pmatrix} 1/5 & 4/5 & 0 & 0 \\ 7/10 & 3/10 & 0 & 0 \\ 0 & 0 & 3/5 & 2/5 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix},$$

where the value appearing in the $(i, j)$-entry of the matrix $\Pi$ is the conditional probability $\pi_{ij} = P\{\tilde{x}_{t+1} = s_j | \tilde{x}_t = s_i\}$ . for all $t$.

$(a)$ Find the stationary probabilistic vector (or vectors, if there are more than one), the ergodic probabilistic vector (if it exists), the transient states (if they exist), the recurrent states (if they exist), and the ergodic sets of the Markov chain.

$(b)$ Compute the matrix $\widehat{\Pi} = \lim_{t\to\infty} \Pi^t$. Find $\lim_{t\to\infty} P\{\tilde{x}_t = s_2 | \tilde{x}_0 = s_2\}$, $\lim_{t\to\infty} P\{\tilde{x}_t = s_1 | \tilde{x}_0 = s_3\}$, and $\lim_{t\to\infty} P\{\tilde{x}_t = s_3 | \tilde{x}_0 = s_4\}$. Assuming that the initial distribution of the Markov chain at $t = 0$ (i.e., the distribution of $\tilde{x}_0$) is such that all the four states are equally likely, what is the unconditional distribution of states in the long run?

11. Consider four different Markov chains on the probability space $(\Omega, \mathcal{F}, P)$ taking values in the state space $S = \{s_1, s_2\}$. These chains are governed by the following four time-invariant transition matrices:

$$\Pi_1 = \begin{pmatrix} 1/2 & 1/2 \\ 1/9 & 8/9 \end{pmatrix}, \; \Pi_2 = \begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \end{pmatrix}, \; \Pi_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \; \Pi_4 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix},$$

where the value appearing in the $(i, j)$-entry of each matrix is the conditional probability $P\{\tilde{x}_{t+1} = s_j | \tilde{x}_t = s_i\}$ for all $t$.

$(a)$ Find the stationary probabilistic vector (or vectors, if there are more than one), the transient states (if they exist), the recurrent states (if they exist), the absorbing states (if they exist) and the ergodic sets for each of the four Markov chains above.

$(b)$ Find the matrices $\widehat{\Pi}_k = \lim_{t\to\infty} \Pi_k^t$, $k = 1, 2, 3, 4$ (if they exist) and the corresponding ergodic probabilistic vectors (if they exist).

$(c)$ We roll a dice to decide the initial state of all those chains. If we get a number of dots smaller or equal than 4, the Markov chains start in state $s_1$; otherwise they start in state $s_2$. Find the unconditional (i.e., before rolling the dice) distributions of states in the long run (if they exist) for the four Markov chains above.

$(d)$ Consider the Markov chain with the transition matrix $\Pi_1$. Find the mean recurrence time for each of the two states of the Markov chain and the mean first passage time from state $s_1$ to $s_2$ and from state $s_2$ to $s_1$.

$(e)$ Consider the two Markov chains with the transition matrices $\Pi_2$ and $\Pi_4$. For these two chains, find the expected number of times the chain is in state $s_1$ given that it starts in state $s_1$. Moreover, for the same two chains, find the expected number of periods before the chain is absorbed when it starts in state $s_1$.

12. Consider a factory with a very large number of workers. All workers have the same very small probability of suffering an accident during a year and all months are equally likely for accidents to occur. We know that the probability of at least one accident occurring in this factory during a year is 0.9, i.e., $P\{\tilde{x} \geq 1\} = 0.9$, where $\tilde{x}$ is the number of accidents in the factory during a year.

    $(a)$ Find the expected number of accidents during a year, $\mathrm{E}(\tilde{x})$.

    $(b)$ Find the probability of three or more accidents occurring during a period of 6 months.

    $(c)$ Suppose now that we know that there have been two or more accidents during a given year. Find the probability that the number $\tilde{x}$ of accidents during this year is larger or equal than one and smaller or equal than three, $P\{1 \leq \tilde{x} \leq 3 | \tilde{x} \geq 2\}$.

13. Let $\tilde{x} \sim \mathrm{N}(0, 1)$ and $\tilde{x}_n = -\tilde{x}$ for $n = 1, 2, 3, ...$ Does $\tilde{x}_n \xrightarrow{d} \tilde{x}$? Does $\tilde{x}_n \xrightarrow{p} \tilde{x}$?

14. Consider a random variable $\tilde{x}$ having the Poisson distribution with parameter $\lambda$.

    $(a)$ Use both the probability function of the Poisson distribution and the Taylor expansion of $e^y$ around $y = 0$ to find the moment generating function $M_{\tilde{x}}(t)$ of the random variable $\tilde{x}$.

($b$) Use the moment generating function found in part (a) to compute the mean $\mu$ and the variance $\sigma^2$ of $\tilde{x}$.

15. Assume that in a given period three firms are created. In every subsequent period, each firm faces the probability of getting bankrupt and disappearing forever. This probability is 1/5 for all firms and for all periods. The mortality of a firm is independent of the other firms' mortality. Find the expected number of periods before all three firms disappear.

16. Consider a Markov chain $\{\tilde{x}_t\}_{t=0}^{\infty}$ on the probability space $(\Omega, \mathcal{F}, P)$. The state space is $S = \{s_1, s_2, s_3\}$ and the time-invariant transition matrix is

$$\Pi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3/5 & 2/5 \\ 0 & 1/2 & 1/2 \end{pmatrix},$$

where the value appearing in the $(i, j)$-entry of the transition matrix $\Pi$ is the conditional probability $\pi_{ij} = P\{\tilde{x}_{t+1} = s_j | \tilde{x}_t = s_i\}$ for all $t$.

($a$) Find the stationary probabilistic vector (or vectors, if there are more than one), the ergodic probabilistic vector (if it exists), the transient states (if they exist), the recurrent states (if they exist), the absorbing states (if they exist), and the ergodic sets of the Markov chain.

($b$) Find the probability in the long run of being in state $s_3$ if the chain starts in state $s_2$,

$$\lim_{t\to\infty} P\{\tilde{x}_t = s_3 | \tilde{x}_0 = s_2\}.$$

($c$) We roll a dice and, if we get either 2 or 6 dots, the Markov chain starts in state $s_1$; if we get an odd number of dots, it starts in state $s_2$; and if we get 4 dots, it starts in state $s_3$. Find the unconditional (i.e., before rolling the dice) probability of being in state $s_2$ in the long run,

$$\lim_{t\to\infty} P\{\tilde{x}_t = s_2\}.$$

17. Show that, if $\tilde{x}$ has a gamma distribution with parameters $\alpha$ and $\beta$, $\tilde{x} \sim \Gamma(\alpha, \beta)$, then the standardization of $\tilde{x}$ converges in distribution to a standard normal random variable when $\alpha \to \infty$ and $\beta$ remains constant. *Hint:* Use Lévy's convergence theorem concerning the equivalence between convergence in distribution and convergence of the corresponding moment-generating function.

18. Use Chebyshev's inequality to prove the weak law of large numbers for a sequence $\{\tilde{x}_i\}_{i=1}^{\infty}$ of independent random variables with identical finite variance, $\text{Var}(\tilde{x}_i) = \sigma^2 < \infty$ for all $i$.