Hence, for each $\epsilon > 0$, $\lim_{n\to\infty} \Pr(A_{n,\epsilon}^C) = 0$, which is precisely what it means to say that $X_n \xrightarrow{P} X$.

Next, assume that $X_n \xrightarrow{P} X$. Let B be a closed set. We shall show that $\limsup_{n\to\infty} \Pr(X_n \in B) \leq \Pr(X \in B)$. Define $C_\epsilon = \{x : d(x, B) \leq \epsilon\}$, a closed set. First note that $\{X_n \in B\} \subseteq \{X \in C_\epsilon\} \cup \{d(X_n, X) > \epsilon\}$. Hence,

$$\Pr(X_n \in B) \leq \Pr(X \in C_\epsilon) + \Pr(d(X_n, X) > \epsilon).$$

Since $X_n \xrightarrow{P} X$, $\lim_{n\to\infty} \Pr(d(X_n, X) > \epsilon) = 0$, so that

$$\limsup_{n\to\infty} \Pr(X_n \in B) \leq \Pr(X \in C_\epsilon),$$

for all $\epsilon > 0$. Since $B = \cup_{m=1}^\infty C_{1/m}$, we know that $\Pr(X \in B) = \lim_{m\to\infty} \Pr(X \in C_{1/m})$, so $\limsup_{n\to\infty} \Pr(X_n \in B) \leq \Pr(X \in B)$. By Theorem B.83 $X_n \xrightarrow{\mathcal{D}} X$.

Next, suppose that X is degenerate at x_0 and $X_n \xrightarrow{\mathcal{D}} X$. Let $\epsilon > 0$, and define

$$g(x) = \begin{cases} 1 & \text{if } d(x, x_0) \leq \frac{\epsilon}{2}, \\ 0 & \text{if } d(x, x_0) \geq \epsilon, \\ 2 - \frac{2d(x,x_0)}{\epsilon} & \text{otherwise.} \end{cases}$$

Since g is bounded and continuous, $\mathrm{E}g(X_n)$ converges to $\mathrm{E}g(X)$. But $\mathrm{E}g(X) = 1$ since $\Pr(g(X) = 1) = 1$, and $\mathrm{E}g(X_n) \leq \Pr(d(X_n, X) < \epsilon)$, since $0 \leq g(x) \leq 1$ for all x. So $\lim_{n\to\infty} \Pr(d(X_n, x_0) < \epsilon) = 1$, and $X_n \xrightarrow{P} X$.

Finally, assume that $X_n \xrightarrow{P} X$. Let n_k be such that $n \geq n_k$ implies

$$\Pr\Big(d(X_n, X) \geq \frac{1}{k}\Big) < 2^{-k}.$$

Define $A_k = \{d(X_{n_k}, X) \geq 1/k\}$. By the first Borel–Cantelli lemma A.20, we have $\Pr(B) = 0$, where $B = \cap_{i=1}^\infty \cup_{k=i}^\infty A_k$. It is easy to check that B is the event that $d(X_{n_k}, X)$ is at least $1/k$ for infinitely many different k. Hence $B^C \subseteq \{\lim_{k\to\infty} X_{n_k} = X\}$, and $\lim_{k\to\infty} X_{n_k} = X$, a.s. $\square$

B.4.2 Characteristic Functions

There is a very important method for proving convergence in distribution which involves the use of *characteristic functions.*

Definition B.91. Let X be a random vector. The complex-valued function

$$\phi_X(t) = \mathrm{E}\left(\exp[it^\top X]\right)$$

is called the *characteristic function of* X. If F is a k-dimensional distribution function, the function $\phi_F(t) = \int \exp[it^\top x] dF(x)$ is called the *characteristic function of* F.