Proofs that I always forget

Andrew Thomas

Exponential potentialities

Bernoulli’s inequality. Let \( x \geq -1 \). Then \((1 + x)^n \geq 1 + nx\).

**Proof.** We proceed by induction. When \( n = 1 \) we have that \( 1 + x \geq 1 + x \). Now suppose that \((1 + x)^n \geq 1 + nx\) holds. Then as \( 1 + x \geq 0 \),
\[
(1 + x)^{n+1} = (1 + x)(1 + x)^n \geq (1 + x)(1 + nx)
= 1 + x + nx + nx^2 = 1 + (n + 1)x + nx^2 \geq 1 + (n + 1)x.
\]

\[\square\]

Exponential limit. Let \( x \in \mathbb{R} \). Then
\[
\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x.
\]

**Proof.** We begin by assuming \( x \geq 0 \). Then, the binomial theorem gives us that
\[
\left(1 + \frac{x}{n}\right)^n = \sum_{k=0}^{n} \binom{n}{k} \frac{x^k}{n^k} = \sum_{k=0}^{n} \frac{x^k}{k!} \frac{n(n-1) \cdots (n-k+1)}{n^k}
\leq \sum_{k=0}^{n} \frac{x^k}{k!} \leq \sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x.
\]

Furthermore, for \( n \geq m \) we have by \( \binom{n}{k} \frac{x^k}{n^k} \geq 0 \) that
\[
\left(1 + \frac{x}{n}\right)^n \geq \sum_{k=m}^{n} \binom{n}{k} \frac{x^k}{n^k} =: s_{n,m}.
\]

Now, as
\[
\frac{n(n-1) \cdots (n-k+1)}{n^k} = \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \to 1,
\]
as \( n \to \infty \) for each fixed \( k \). Hence, \( \liminf_{n \to \infty} s_{n,m} = \sum_{k=0}^{m} \frac{x^k}{k!} \) so we conclude
\[
e^x = \liminf_{m \to \infty} \left( \liminf_{n \to \infty} s_{n,m} \right) \leq \liminf_{n \to \infty} \left(1 + \frac{x}{n}\right)^n \leq \limsup_{n \to \infty} \left(1 + \frac{x}{n}\right)^n \leq e^x.
\]

For large enough \( n \), \(|x/n| \leq 1\) so that
\[
\left(1 - \frac{x}{n}\right)^n \left(1 + \frac{x}{n}\right)^n = \left(1 - \frac{x^2}{n^2}\right)^n \leq 1,
\]
but we also know by Bernoulli’s inequality
\[
\left(1 - \frac{x^2}{n^2}\right)^n \geq 1 - \frac{x^2}{n}.
\]
Taking \(\lim\inf\) and \(\lim\sup\) we get
\[
\lim_{n \to \infty} \left( 1 - \frac{x}{n} \right)^n \left( 1 + \frac{x}{n} \right)^n = 1,
\]
implying that
\[
\lim_{n \to \infty} \left( 1 - \frac{x}{n} \right)^n = \lim_{n \to \infty} \frac{ \left( 1 - \frac{x^2}{n^2} \right)^n } { \left( 1 + \frac{x}{n} \right)^n } = e^{-x}.
\]

A trivial corollary as a result of the above proof is that for any \(n\) we have \((1 + x/n)^n \le e^x\).

**An interesting logarithmic inequality.** For all \(x > 0\) we have that
\[
\log(1 + 1/x) \ge \frac{1}{1 + x}
\]

**Proof.** First note that for \(x > 0\) we have that
\[
\frac{1}{(x+1)^2} \le \frac{1}{x(x+1)},
\]
thus we have
\[
\int_x^\infty \frac{1}{(t+1)^2} \, dt \le \int_x^\infty \frac{1}{t(t+1)} \, dt = \lim_{y \to \infty} \int_x^y \frac{1}{t} - \frac{1}{t+1} \, dt.
\]

Now, as
\[
\lim_{y \to \infty} \int_x^y \frac{1}{t} - \frac{1}{t+1} \, dt = \lim_{y \to \infty} -\log(1 + 1/t)|_x^y = \log(1 + 1/x),
\]
and
\[
\int_x^\infty \frac{1}{(t+1)^2} \, dt = \frac{1}{1 + x},
\]
the proof is finished.

As a rather interesting corollary to this we can establish that \(e \le (1 + 1/x)^{x+1}\) for any \(x > 0\) – in particular the natural numbers. This is equivalent to showing that \((x+1) \log(1+1/x) \ge 1\) or rather that \(\log(1+1/x) \ge 1/(1 + x)\), which is exactly the result from above.

**A technical result for the DeMoivre-Laplace theorem.** Let \(0 \le p \le 1\). Furthermore, suppose that for \(k_n(x)\) is a sequence of integers such that \(|k_n(x)| \le n\) and
\[
\frac{k_n(x) - np}{\sqrt{np(1-p)}} \to x,
\]
where \(x \in \mathbb{R}\). Abbreviating \(k_n := k_n(x)\), we have that
\[
\left( \frac{np}{k_n} \right)^{k_n} \left( \frac{n(1-p)}{n-k_n} \right)^{n-k_n} \to e^{-x^2/2},
\]
as \(n \to \infty\).

**Proof.** We will prove that
\[
k_n \log \left( \frac{np}{k_n} \right) + (n - k_n) \log \left( \frac{n(1-p)}{n-k_n} \right) \to -\frac{x^2}{2},
\]
which implies our result as \(x \to e^x\) a continuous function. If we recognize that
\[
\frac{np}{k_n} = 1 - \frac{k_n - np}{k_n} \quad \text{and} \quad \frac{n(1-p)}{n-k_n} = 1 + \frac{k_n - np}{n-k_n},
\]

Thus we have that as we will now show. This convergence follows as a result of the fact that

\[ \log(1 + x) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n}. \] 

Thus we have that

\[ \log \left( \frac{np}{k_n} \right) = -\frac{(k_n - np)}{k_n} - \frac{1}{2} \left( \frac{k_n - np}{k_n} \right)^2 + O(n^{-3/2}), \]

and

\[ \log \left( \frac{n(1 - p)}{n - k_n} \right) = -\frac{(k_n - np)}{n - k_n} - \frac{1}{2} \left( \frac{k_n - np}{n - k_n} \right)^2 + O(n^{-3/2}). \]

Hence, we arrive at

\[
k_n \log \left( \frac{np}{k_n} \right) + (n - k_n) \log \left( \frac{n(1 - p)}{n - k_n} \right) = k_n \left[ -\frac{(k_n - np)}{k_n} - \frac{1}{2} \left( \frac{k_n - np}{k_n} \right)^2 + O(n^{-3/2}) \right] + (n - k_n) \left[ -\frac{(k_n - np)}{n - k_n} - \frac{1}{2} \left( \frac{k_n - np}{n - k_n} \right)^2 + O(n^{-3/2}) \right]
\]

\[= -\frac{(k_n - np)^2}{2k_n} - \frac{(k_n - np)^2}{2(n - k_n)} + O(n^{-1/2}) \to -\frac{x^2}{2}, \]

as we will now show. This convergence follows as a result of the fact that

\[-\frac{(k_n - np)^2}{2k_n} = -\frac{np(1-p)}{k_n} \left( \frac{(k_n - np)^2}{2np(1-p)} \right) \to -\frac{x^2(1-p)}{2}, \]

as \( x \mapsto -x^2 \) is continuous and \( k_n/np \to 1 \). The other case follows similarly. To demonstrate \( k_n/np \to 1 \) just take

\[ \frac{k_n}{np} = 1 + \frac{k_n - np}{np} = 1 + \frac{1-p}{np} \frac{k_n - np}{\sqrt{np(1-p)}} \to 1 + 0 \cdot x = 1. \]

\[ \square \]

**Expo-linear inequalities.** Let \( x \in \mathbb{R} \). Then \( 1 + x \leq e^x \). For \( x \geq 0 \), we have \( e^{-x^2/2} \leq e^x (1 + x)^{-(1+x)} \).

**Proof.** If \( y \geq 0 \), then \( 1 \geq 1/(1+y) \) hence

\[ \log(1 + x) = \int_0^x \frac{1}{1+y} \, dy \leq \int_0^x 1 \, dy = x, \]

though this is also obvious due to the series representation of \( e^x \). For \( x \leq 0 \), let \( z := -x \) and let us show that \( 1 - z \leq e^{-z} \), or \( 1 - e^{-z} \leq z \) for \( z \geq 0 \). This follows, as

\[ 1 - e^{-z} = \int_0^z e^{-y} \, dy \leq \int_0^z 1 \, dy = z. \]

We integrate again using our first inequality to get

\[ \int_0^x \log(1 + y) \, dy \leq \int_0^x y \, dy. \]

As

\[ \int_0^x \log(1 + y) \, dy = x \log(1 + x) - \int_0^x \frac{y}{1+y} \, dy = x \log(1 + x) - x + \log(1 + x) = (1 + x) \log(1 + x) - x, \]

we get that \( (1 + x) \log(1 + x) \leq x + x^2/2 \) or \( (1 + x)^{1+x} \leq e^{e^x/2} \), so that we get \( e^{-x^2/2} (1 + x)^{1+x} \leq e^x \) or

\[ e^{-x^2/2} \leq e^x (1 + x)^{-(1+x)}. \]

\[ \square \]
Though, as a final note for \( x \geq 0 \), it is rather obvious that our first “expo-linear” inequality holds as \( e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \geq \sum_{k=0}^{n} \frac{x^k}{k!} \) for any \( n \in \mathbb{N} \).

**Binomial power bounds.** If \( n, m, k \) non-negative integers such that \( n \geq k \) then we have that

\[
\binom{n + m}{k + 1} - \binom{n}{k + 1} \leq \frac{m(n + m)^k}{k!}
\]

*Proof.* We will proceed by induction. First note that if \( k = 0 \) then we have equality. If \( m = 1 \) and \( k \geq 2 \) then

\[
\binom{n + 1}{k + 1} - \binom{n}{k + 1} = \frac{(n + 1) \cdots (n + 1 - k) - n(n - 1) \cdots (n - k)}{(k + 1)!}
\]

\[
= \frac{n(n - 1) \cdots (n - k + 1)(n + 1 - (n - k))}{(k + 1)!}
\]

\[
= \binom{n}{k} \leq \frac{n^k}{k!}.
\]

Now if we assume the desired inequality holds for \( m \), then

\[
\binom{n + m + 1}{k + 1} - \binom{n}{k + 1} = \binom{n + m + 1}{k + 1} - \binom{n + m}{k + 1} + \binom{n + m}{k + 1} - \binom{n}{k + 1}
\]

\[
\leq \binom{n + m}{k} + \frac{m(n + m)^k}{k!}
\]

\[
\leq \frac{(m + 1)(n + m + 1)^k}{k!},
\]

hence proved. \( \square \)

**Topological curiosities**

**Continuous images of compact sets are compact.** Let \( f : X \to Y \) be a continuous function from the space \( X \) into the space \( Y \). If \( C \subset X \) is compact, then \( f(C) \) is compact.

*Proof.* If \( \{V_\alpha\}_\alpha \) is an open cover of \( f(C) \), and if \( U_\alpha = f^{-1}(V_\alpha) \), then \( \{U_\alpha\}_\alpha \) an open cover of \( C \). Let \( U_{\alpha_1}, \ldots, U_{\alpha_n} \) be a finite subcover of \( C \), that is

\[ C \subset \bigcup_{\alpha=1}^{\alpha_n} U_\alpha = f^{-1}(\bigcup_{\alpha=1}^{\alpha_n} V_\alpha), \]

which implies that if \( x \in C \), then \( f(x) \in \bigcup_{\alpha=1}^{\alpha_n} V_\alpha \). Then as \( y = f(x) \in f(C) \) for some \( x \in C \) by definition this entails that \( f(C) \subset \bigcup_{\alpha=1}^{\alpha_n} V_\alpha \) and hence \( f(C) \) compact. \( \square \)

**Product of the subspace topology is the same as the subspace topology of a product.** Suppose that \( Y_\alpha \) is a subspace of \( X_\alpha \), for \( \alpha \in I \) for some index set \( I \). Endow \( Y := \prod_{\alpha \in I} Y_\alpha \) and \( X := \prod_{\alpha \in I} X_\alpha \) with their respective product topologies. Then \( Y \) is a subspace of \( X \) with respect to the product topology of \( X \).

*Proof.* First, let \( \mathcal{T}_\alpha \) be the topology of \( X_\alpha \) and let \( \mathcal{T}'_\alpha \) be the subspace topology of \( Y_\alpha \), that is \( \mathcal{T}'_\alpha = \{U \cap Y_\alpha : U \in \mathcal{T}_\alpha \} \). Now, for any \( k \in \mathbb{N} \) and \( \alpha_1, \ldots, \alpha_k \in I \) define the projection mapping \( \pi_{\alpha_1, \ldots, \alpha_k} : X \to X_{\alpha_1} \times \cdots \times X_{\alpha_k} \) by \( \pi_{\alpha_1, \ldots, \alpha_k}(x) = (x_{\alpha_1}, \ldots, x_{\alpha_k}) \). Let \( \pi'_{\alpha_1, \ldots, \alpha_k} \) be the corresponding projection for \( Y \). Let us take \( x := (x_\alpha)_{\alpha \in I} \) to be a typical element of \( X \). Let \( \mathcal{T}' \) be the topology with the basis elements

\[
(\pi'_{\alpha_1, \ldots, \alpha_k})^{-1}(U'_{\alpha_1} \times \cdots \times U'_{\alpha_k}),
\]
where $U'_{\alpha_i} \in \mathcal{T}'_{\alpha_i}$ for each $i = 1, \ldots, k$. That is, the product topology of the $Y_{\alpha}$ subspaces. Now let $\mathcal{T}''$ be the subspace topology of $Y$ in $X$. That is,
$$\mathcal{T}'' = \{U \cap Y : U \in \mathcal{T}\},$$
which has basis elements $\pi_{\alpha_1,\ldots,\alpha_k}^{-1}(U_{\alpha_1} \times \cdots \times U_{\alpha_k}) \cap Y$ for $U_{\alpha_i} \in \mathcal{T}_{\alpha_i}$. The goal is to show that $\mathcal{T}' = \mathcal{T}''$.

Suppose that $y \in Y$ and let $B \in \mathcal{T}'$ be a basis element containing $y$. Then $y \in (\pi_{\alpha_1,\ldots,\alpha_k}^{-1}(U'_{\alpha_1} \times \cdots \times U'_{\alpha_k})$ for some $\alpha_1, \ldots, \alpha_k \in I$. However, for each $i$, $U'_{\alpha_i} = U_{\alpha_i} \cap Y_{\alpha_i}$ for some $U_{\alpha_i} \in \mathcal{T}_{\alpha_i}$ and thus,
$$\begin{align*}
&\left(\pi_{\alpha_1,\ldots,\alpha_k}^{-1}(U'_{\alpha_1} \times \cdots \times U'_{\alpha_k})
\right) \\
&\quad = \left\{ x \in Y : \pi'_{\alpha_1,\ldots,\alpha_k}(x) \in \prod_{i=1}^{k}(U_{\alpha_i} \cap Y_{\alpha_i}) \right\} \\
&\quad = \left\{ x \in Y : \pi'_{\alpha_1,\ldots,\alpha_k}(x) \in \prod_{i=1}^{k}U_{\alpha_i} \cap \prod_{i=1}^{k}Y_{\alpha_i} \right\} \\
&\quad = Y \cap \left\{ x \in X : \pi_{\alpha_1,\ldots,\alpha_k}(x) \in \prod_{i=1}^{k}U_{\alpha_i} \right\} \\
&\quad = Y \cap \pi_{\alpha_1,\ldots,\alpha_k}(U_{\alpha_1} \times \cdots \times U_{\alpha_k}) \in \mathcal{T}'',
\end{align*}$$
and $y$ obviously contained in the latter set. Symmetry furnishes us with the other inclusion. Hence we have that $\mathcal{T}' = \mathcal{T}''$. □

This theorem gives a useful corollary to Tychonoff’s theorem. Namely, if $K_i$ are compact subspaces of $X_i$, then $\prod_{i \in I} K_i$ is compact when endowed with the product-of-subspace topology. The above implies that it is compact when endowed with the subspace-of-product topology, as a subspace of $\prod_{i \in I} X_i$. Hence $\prod_{i \in I} K_i$ is compact in the sense that every open cover of sets in $\prod_{i \in I} X_i$ has a finite subcover.

As it concerns probability, this implies that any cartesian product of probability measure is tight if each of marginal measures are tight.

**Unaccompanied lemmas**

**Limit infimum and indicators.** Let $\{A_n\}_{n \geq 1}$ be subsets of a set $X$. Then for every $x \in X$ we have
$$1_{\lim\inf_{n \to \infty} A_n}(x) = \lim\inf_{n \to \infty} 1_{A_n}(x),$$
and the corresponding fact holds for the limit supremum.

**Proof.** We first mention that if $s_n \in \mathbb{R}$ and $s = \lim inf s_n$ then for every $y < s$ there exists an $N_y$ such that $s_n > y$ for $n \geq N_y$. Now, suppose that $\lim inf_{n \to \infty} 1_{A_n}(x) = 1$. Then by the aforementioned property of the limit infimum, we have that there exists an $N$ such that if $n \geq N$ then $1_{A_n}(x) = 1$, as our sequence can only take the values 0 or 1. Hence, $x \in \lim inf A_n$ so that $1_{\lim inf_{n \to \infty} A_n}(x) = 1$.

Now suppose that $1_{\lim inf_{n \to \infty} A_n}(x) = 1$, then by definition of limit infimum for sets, there exists some $N$ such that for $n \geq N$ we have that $x \in A_n$ which is equivalent to $1_{A_n}(x) = 1$. Hence, $1_{A_n}(x) \to 1$ and hence $\lim inf_{n \to \infty} 1_{A_n}(x) = 1$. Thus
$$1_{\lim inf_{n \to \infty} A_n}(x) = 1 \iff \lim inf_{n \to \infty} 1_{A_n}(x) = 1,$$
and clearly the cases when either of these are zero are equivalent as well. □

**Double convergence.** Let $f_n : X \to Y$ be a sequence of continuous functions from a space $X$ to a metric space $(Y,d)$. Suppose that $f_n \to f$ uniformly and that $x_n \to x$. Then $f_n(x_n) \to f(x)$.
Proof. Let $\epsilon > 0$ be given. By the uniform limit theorem, we have that $f$ is continuous. Hence we can find an $N_1$ such that if $n \geq N_1$ then $d(f(x), f(x_n)) < \epsilon/2$. Additionally, as $\{f_n\}$ converges to $f$ uniformly then we can find an $N_2$ such that if $n \geq N_2$ then $d(f_n(t), f(t)) < \epsilon/2$ for all $t \in X$. So if $n \geq N_1 \lor N_2$, then
\[
d(f_n(x_n), f(x)) \leq d(f(x), f(x_n)) + d(f(x_n), f_n(x_n)) < \epsilon/2 + \epsilon/2 = \epsilon.
\]
\[\square\]

**Remark 1.** The uniformity assumption is essential. Let us take $f_n : [0,1] \rightarrow [0,1]$ be defined by $f_n(x) = x^n$. Then for each $x \in [0,1]$ we have
\[
\lim_{n \rightarrow \infty} f_n(x) \rightarrow \delta_1(x),
\]
the dirac delta function defined on the unit interval. Now, if we take $x_n = 1 - \frac{1}{n}$ then clearly $x_n \rightarrow 1$ as $n \rightarrow \infty$. However,
\[
\lim_{n \rightarrow \infty} f_n(x_n) = \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right)^n = e^{-1} \neq 1.
\]
Note that
\[
\sup_{x \in [0,1]} |f_n(x) - \delta_1(x)| = 1,
\]
for all $n$, so does not converge to $0$.

**A canonical outer measure for the Carathéodory’s Extension Theorem.** Let $\mu_0$ be a pre-measure on an algebra $\mathcal{A}$ in $X$. Define $s : \mathcal{P}(X) \rightarrow \mathcal{P}(\mathbb{R})$ as
\[
s(E) := \left\{ \sum_{n=1}^{\infty} \mu_0(E_n) : E \subset \bigcup_{n=1}^{\infty} E_n, E_n \in \mathcal{A} \text{ for all } n \right\}
\]
Then we have that $\mu^*$ defined by $\mu^*(E) := \inf s(E)$ for all $E \subset X$ is an outer measure.

**Proof.** First we note that $\emptyset \subset \emptyset$ and so $\mu^*(\emptyset) = \mu_0(\emptyset) = 0$. We now aim to show monotonicity. If $E \subset F \subset \bigcup_{n=1}^{\infty} F_n$ with $F_n \in \mathcal{A}$ then we have that $s(F) \subset s(E)$ which implies $\mu^*(E) \leq \mu^*(F)$ as if $A \subset B$ are subsets of $[-\infty, \infty]$ then we have that inf $A \geq$ inf $B$.

Finally, we must show countable subadditivity. By definition of infimum for any $\epsilon > 0$ we can find a sequence $E_{n,1}, E_{n,2}, \ldots$ of sets in $\mathcal{A}$ such that $\mu^*(E_{n,1}) + \epsilon 2^{-n} > \sum_{m=1}^{\infty} \mu_0(E_{n,m})$, with $E_n \subset \bigcup_{m=1}^{\infty} E_{n,m}$. Thus, it is clear that
\[
\bigcup_{n=1}^{\infty} E_n \subset \bigcup_{(n,m) \in \mathbb{N}^2} E_{n,m},
\]
and as the right-hand side is a countable union of sets in $\mathcal{A}$, invoking Tonelli’s theorem we have
\[
\mu^* \left( \bigcup_{n=1}^{\infty} E_n \right) \leq \sum_{(n,m) \in \mathbb{N}^2} \mu_0(E_{n,m}) \leq \sum_{n=1}^{\infty} \mu^*(E_n) + \epsilon,
\]
As $\epsilon$ is arbitrary we have our result. \[\square\]

**Uniform continuity and modulus.** Let $f : X \rightarrow Y$ be a continuous function from a metric space $(X, d_X)$ to another metric space $(Y, d_Y)$. Define
\[
\omega_\delta(f) := \sup \{d_Y(f(s), f(t)) : d_X(s,t) \leq \delta\},
\]
to be the modulus of continuity of $f$. Then $f$ is uniformly continuous if and only if $\omega_\delta(f) \rightarrow 0$ as $\delta \rightarrow 0$.

**Proof.** Suppose that $f$ is uniformly continuous. Let us take $\delta_n \rightarrow 0$ and suppose that $\omega_{\delta_n}(f) \not\rightarrow 0$. We proceed by the following steps:
1. There exists an \( \epsilon > 0 \) such that for all \( N \), there exists an \( n \geq N \) such that \( \omega_{\delta_n}(f) \geq \epsilon \).

2. Now, by uniform continuity there exists a \( \delta > 0 \) such that if \( d_X(s, t) < \delta \), then \( d_Y(f(s), f(t)) < \epsilon/2 \).

3. By convergence of \( \{\delta_n\} \) there exists an \( N_\delta \) such that if \( n \geq N_\delta \) then \( \delta_n < \delta \).

4. Finally, there exists an \( n_0 \geq N_\delta \) such that \( \omega_{\delta_{n_0}}(f) \geq \epsilon \).

However, as \( d_X(s, t) \leq \delta_{n_0} < \delta \) then \( \omega_{\delta_{n_0}} \leq \epsilon/2 \), a contradiction.

Now suppose that \( \omega_\delta(f) \to 0 \) as \( \delta \to 0 \). Then for every \( \epsilon > 0 \) there exists an \( \eta > 0 \) such that if \( \delta < \eta \) then \( \omega_\delta(f) < \epsilon \). Now if \( d_X(s, t) \leq \delta \) then \( d_Y(f(s), f(t)) \leq \omega_\delta(f) < \epsilon \). Hence, \( f \) is uniformly continuous.

\[\square\]

**Measures and the “standard” machine**

**Pointwise limits of measurable functions are measurable.** Let \( (X, \mathcal{B}) \) be a measurable space. Suppose that \( f_n : X \to [0, \infty] \) are a sequence of measurable functions that convergence pointwise to a limit \( f : X \to [0, \infty] \). Show that \( f \) is measurable.

**Proof.** For \( f \) to be measurable it is necessary and sufficient that

\[\{x \in X : f(x) \geq a\} \in \mathcal{B},\]

for every \( a \geq 0 \), by theorem 1.3.1 in [3].

By our measurability hypothesis, we have that for every \( n \in \mathbb{N} \) and \( a \geq 0 \) that \( \{x \in X : f_n(x) \geq a\} \in \mathcal{B} \).

Suppose that \( x \) is such that \( f(x) \geq a \). Then for any \( \epsilon > 0 \) there exists an \( N \) such that for all \( n \geq N \) we have \( a \leq f(x) \leq f_n(x) + \epsilon \). Hence for every \( m \in \mathbb{N} \),

\[\{x \in X : f(x) \geq a\} \subset \bigcap_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ x \in X : f_n(x) \geq a - \frac{1}{m} \right\}.\]

Now, suppose that \( x \) is such that \( f(x) < a \), and hence \( f(x) < a - 1/m \) for some \( m \in \mathbb{N} \). Then there exists some \( m \) such that for every \( N \) we can find an \( n \geq N \) such that \( f_n(x) < a - 1/m \). Hence the two sets are equal and the measurability of \( f \) follows from the fact that \( \mathcal{B} \) a \( \sigma \)-algebra.

\[\square\]

**Integral with respect to Dirac measure.** Let \( f : X \to \mathbb{C} \) be a complex-valued function on the measurable space \( (X, \mathcal{B}) \). Assume that \( f \) is \( \mathcal{B} \)-measurable. Then if \( x \in X \) we have that

\[\int_X f(y) \delta_x(dy) = f(x)\]

**Proof.** Let \( A \in \mathcal{B} \). Then by definition of an integral of a simple function with respect to an abstract measure, we have that

\[\int_X 1_A(y) \delta_x(dy) = \delta_x(A) = 1_A(x),\]

as desired. Similarly, we have for a simple function \( \sum_{i=1}^{m} c_i 1_{A_i} \) with non-negative constants \( c_1, \ldots, c_m \) that

\[\int_X \sum_{i=1}^{m} c_i 1_{A_i}(y) \delta_x(dy) = \sum_{i=1}^{m} c_i \delta_x(A_i) = \sum_{i=1}^{m} c_i 1_{A_i}(x),\]

again by the aforementioned properties for integrals of simple functions. Now for any \( \mathcal{B} \)-measurable \( f \geq 0 \) there exist simple functions \( f_n \) such that \( f_n \uparrow f \) pointwise. Then, by monotone convergence we have

\[f_n(x) = \int_X f_n(y) \delta_x(dy) \uparrow \int_X f(y) \delta_x(dy),\]

which implies \( \int_X f(y) \delta_x(dy) = f(x) \) by the uniqueness of limits in \( \mathbb{R} \). The final parts follow from the fact we can break a real (complex) function into its positive and negative (real and imaginary) parts.

\[\square\]
Independence with expectations. Let \( X_1, X_2, \ldots, X_k \) be random elements of a measurable space \((X, \mathcal{B})\) defined on some probability space \((\Omega, \mathcal{F}, P)\). Then \( X_1, X_2, \ldots, X_k \) independent if and only if for every \( f_1, \ldots, f_k : X \to \mathbb{R} \) bounded and \( \mathcal{B} \)-measurable we have

\[
E \left[ \prod_{i=1}^{k} f_i(X_i) \right] = \prod_{i=1}^{k} E \left[ f_i(X_i) \right]
\]

Proof. As an aside, we note that

An infinite collection of random elements is independent if every finite subcollection is independent.

If we let \( f_i := 1_{A_i} \) for some \( A_i \in \mathcal{B} \) then we have

\[
E \left[ \prod_{i=1}^{k} 1_{A_i}(X_i) \right] = P(\cap_{i=1}^{k} \{ X_i \in A_i \}) = \prod_{i=1}^{k} P(X_i \in A_i) = \prod_{i=1}^{k} E \left[ 1_{A_i}(X_i) \right].
\]

Now let us suppose that \( X_1, X_2, \ldots, X_k \) are independent. Then clearly for indicator functions \( f_i \) as just described we have that the expectations factor. We proceed now to simple functions \( f_i := \sum_{j=1}^{n_i} c_{i,j} 1_{A_{i,j}} \), where \( A_{i,j} \in \mathcal{B} \), and \( c_{i,j} \geq 0 \). First let \( J_i := \{1, \ldots, n_i\} \) and let \( J := J_1 \times J_2 \times \cdots \times J_k \).

\[
E \left[ \prod_{i=1}^{k} f_i(X_i) \right] = E \left[ \prod_{i=1}^{k} \left( \sum_{j=1}^{n_i} c_{i,j} 1_{A_{i,j}}(X_i) \right) \right] = \sum_{(j_1, j_2, \ldots, j_k) \in J} \prod_{i=1}^{k} c_{i,j_i} E \left[ 1_{A_{i,j_i}}(X_i) \right]
\]

Generalizing to non-negative functions, let us take for each \( i \) a sequence of simple functions \( f_{i,n} \) as previously described, such that \( f_{i,n} \uparrow f_i \), with \( f_i : X \to [0, \infty) \) bounded \( \mathcal{B} \)-measurable functions. Boundedness guarantees all our expectations exist. For each \( n \) we have that

\[
E \left[ \prod_{i=1}^{k} f_{i,n}(X_i) \right] = \prod_{i=1}^{k} E \left[ f_{i,n}(X_i) \right].
\]

However, the monotone convergence theorem and the fact that the product of limits is the limit of the products (which entails \( \prod_{i=1}^{k} f_{i,n} \uparrow \prod_{i=1}^{k} f_i \)), gives us our result. Finally, suppose that for each \( i \) we have that \( f_i : X \to \mathbb{R} \) are \( \mathcal{B} \)-measurable and bounded. Then \( f_i = f_i^+ - f_i^- \) and each of these functions are bounded. Now,

\[
E \left[ \prod_{i=1}^{k} f_i(X_i) \right] = \prod_{i=1}^{k} E \left[ f_i(X_i) \right],
\]

comes as result that \( \prod_{i=1}^{k} f_i^\pm \) is a bounded, non-negative measurable function. \( \square \)

Linear change of variable – Lebesgue integral. Let \( f : \mathbb{R}^d \to [0, \infty] \) and \( T : \mathbb{R}^d \to \mathbb{R}^d \) be an invertible linear transformation. Then

\[
\int_{\mathbb{R}^d} f(T^{-1}(x)) \, dx = |\det(T)| \int_{\mathbb{R}^d} f(x) \, dx
\]

Proof. We begin by noting the following result:

If \( A \subset \mathbb{R}^d \) is measurable, then so is \( T(A) \subset \mathbb{R}^d \) and \( m(T(A)) = |\det(T)|m(A) \),
where $m$ is $d$-dimensional Lebesgue measure (cf. [1]). Suppose that $T^{-1}(x) \in A$ for some $x \in \mathbb{R}^d$. By $T$ invertible, it is bijective and hence $x \in T(A)$. If $x \in T(A)$, then $x = T(y)$ for some $y \in A$ and hence $T^{-1}(x) = T^{-1}(T(y)) = y \in A$. The other direction is fairly obvious. Hence, if $f := 1_A$, we get

$$\int_{\mathbb{R}^d} 1_A(T^{-1}(x)) \, dx = \int_{\mathbb{R}^d} 1_{T(A)}(x) \, dx = m(T(A)) = |\det(T)| m(A),$$

by what was mentioned at the beginning of the proof. Now suppose that $f := \sum_{i=1}^n c_i 1_{A_i}$, where $A_i$ measurable and $c_i \geq 0$ and $m(A_i) < \infty$ for all $i$. That is, $f$ a non-negative bounded simple function with finite measure support. Then

$$\int_{\mathbb{R}^d} f(T^{-1}(x)) \, dx = \int_{\mathbb{R}^d} \sum_{i=1}^n c_i 1_{A_i}(T^{-1}(x)) \, dx = \sum_{i=1}^n c_i \int_{\mathbb{R}^d} 1_{T(A_i)}(x) \, dx = |\det(T)| \sum_{i=1}^n c_i 1_{A_i}(x) \, dx = |\det(T)| \int_{\mathbb{R}^d} f(x) \, dx.$$

Now if $f : \mathbb{R}^d \to [0, \infty]$ is measurable, then there exist non-negative bounded simple functions with finite measure support $f_n$ such that $f_n \uparrow f$ pointwise [1]. The monotone convergence theorem and the equality just established furnishes our final result.

\[\square\]

The probabilistic touch

**Separating class theorem.** Let probability measures $P$ and $Q$ be defined on the measurable space $(\Omega, \mathcal{F})$. Suppose that $P(A) = Q(A)$ for all $A \in \mathcal{P}$, where $\mathcal{P}$ a $\pi$-system. Then $P(A) = Q(A)$ for all $A \in \sigma(\mathcal{P})$. If $\mathcal{F} = \sigma(\mathcal{C})$ for some collection $\mathcal{C}$ of sets, then $P(A) = Q(A)$ for all $A \in \mathcal{F}$ if $\mathcal{C} \subset \sigma(\mathcal{P})$.

**Proof.** Begin by denoting

$$\mathcal{L} = \{ A \in \sigma(\mathcal{P}) : P(A) = Q(A)\}.$$

Clearly we have that $\mathcal{L} \subset \sigma(\mathcal{P})$. If we can show that $\mathcal{L}$ a $\lambda$-system, then by the $\pi - \lambda$ theorem we have that $\mathcal{P} \subset \mathcal{L}$ implies that $\sigma(\mathcal{P}) \subset \mathcal{L}$. By the fact that $P(\Omega) = Q(\Omega) = 1$, $\Omega \in \mathcal{L}$. Now suppose that $A \in \mathcal{L}$. Then one sees that $P(A^c) = 1 - P(A) = 1 - Q(A) = Q(A^c)$ so $A^c \in \mathcal{L}$. Now take a sequence of disjoint sets $A_1, A_2, \cdots \in \mathcal{L}$. We observe by countable additivity that

$$P(\cup_{i} A_i) = \sum_{i} P(A_i) = \sum_{i} Q(A_i) = Q(\cup_{i} A_i),$$

because identical sequences must have the same limit – or rather $\lim_{n \to \infty} \sum_{i=1}^n P(A_i) = \lim_{n \to \infty} \sum_{i=1}^n Q(A_i)$. Therefore, $\mathcal{L}$ is a $\lambda$-system and thus $\sigma(\mathcal{P}) \subset \mathcal{L}$ so $\sigma(\mathcal{P}) = \mathcal{L}$ and $P$ and $Q$ coincide on $\sigma(\mathcal{P})$.

For the additional condition stated above, we see that as $\sigma(\mathcal{C})$ minimal, then $\sigma(\mathcal{C}) \subset \sigma(\mathcal{P})$ and so $P$ and $Q$ coincide on $\mathcal{F}$. \[\square\]

**Countable totality.** Let us take the probability space $(\Omega, \mathcal{F}, P)$ and let $\{A_\alpha\}$ be an uncountable family of disjoint subsets of $\Omega$ measurable with respect to $\mathcal{F}$. Then at most only countably many sets in $\{A_\alpha\}$ have positive probability.

**Proof.** Let $A_n := \{ \alpha : P(A_\alpha) \geq \frac{1}{n} \}$. This implies that $|A_n| \leq n$ as otherwise if $|A_n| > n$ for some $N$ then

$$P\left( \bigcup_{\alpha \in A_n} A_\alpha \right) = \sum_{\alpha \in A_n} P(A_\alpha) > 1,$$

a contradiction. Now we aim to show that $A := \{ \alpha : P(A_\alpha) > 0 \} = \cup_n A_n$. The latter set is clear contained in the first. Now suppose that $P(A_n) > 0$. Then there exists an $\epsilon > 0$ such that $P(A_\alpha) > \epsilon$ and an $N$ such that $1/n < \epsilon$ for $n \geq N$. Thus $\alpha \in A_N$ so we conclude that $A = \cup_n A_n$. Hence, as each $A_n$ finite then we have that $A$ at most countable. \[\square\]
Measurability preserves independence. Let \( X \) and \( Y \) be independent random variables defined on a probability space \((\Omega, \mathcal{F}, P)\) and let \( f, g \) be measurable functions. Then \( f(X), g(Y) \) are independent.

Proof. Independence implies that for every \( A, B \in \mathcal{B}(\mathbb{R}) \), the Borel \( \sigma \)-algebra on the real line, that

\[
P(X \in A, Y \in B) = P(X \in A)P(Y \in B).
\]

If we note that as \( f, g : \mathbb{R} \to \mathbb{R} \) are measurable this implies for every \( C \in \mathcal{B}(\mathbb{R}) \) that \( f^{-1}(C), g^{-1}(C) \in \mathcal{B}(\mathbb{R}) \).

Now, suppose that \( \omega \in \{ f(X) \in A \} \), this is the same as saying \( f(X(\omega)) \in A \), which is equivalent to the statement \( X(\omega) \in f^{-1}(A) \). Thus \( \{ f(X) \in A \} = \{ X \in f^{-1}(A) \} \). Using the same argument for \( g(Y) \) we have for every \( A, B \in \mathcal{B}(\mathbb{R}) \) that

\[
P(f(X) \in A, g(Y) \in B) = P(f(X) \in f^{-1}(A), Y \in g^{-1}(B))
\]

\[
= P(f(X) \in f^{-1}(A))P(Y \in g^{-1}(B))
\]

\[
= P(f(X) \in A)P(g(Y) \in B).
\]

\( \square \)

Correlation is less than 1 in absolute value. Let \( X, Y \in L^2 \). Then if we define

\[
\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}},
\]

then \( |\rho_{X,Y}| \leq 1 \).

Proof. Suppose first that \( E[X] = E[Y] = 0 \). Then Jensen’s inequality and Cauchy-Schwarz implies that

\[
|\rho_{X,Y}| = \frac{|E[XY]|}{\sqrt{E[X^2]E[Y^2]}} \leq \frac{E|XY|}{\sqrt{E[X^2]E[Y^2]}} \leq \sqrt{\frac{E[X^2]E[Y^2]}{E[X^2]E[Y^2]}} \leq 1.
\]

Now if \( E[X] = \mu_X \) and \( E[Y] = \mu_Y \), and we let \( X' := X - \mu_X \) and \( Y' := Y - \mu_Y \) then we have \( E[X'] = E[Y'] = 0 \). If we note the equality

\[
\rho_{X',Y'} = \frac{E[XY - \mu_X Y - X \mu_Y + \mu_X \mu_Y]}{\sqrt{E[(X - \mu_X)^2]E[(Y - \mu_Y)^2]}} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \rho_{X,Y},
\]

and that \( |\rho_{X',Y'}| \leq 1 \) from above, this implies that \( |\rho_{X,Y}| \leq 1 \).

\( \square \)

Integrate out to get conditional expectation. Let \( X_1, \ldots, X_k, X_{k+1}, \ldots, X_n \) be independent random variables defined on \((\Omega, \mathcal{F}, P)\), where \( X_i \) has the distribution \( F_i \). Let us define measurable functions \( \phi : \mathbb{R}^n \to \mathbb{R} \) and \( g : \mathbb{R}^k \to \mathbb{R} \) by

\[
g(x_1, \ldots, x_k) = E[\phi(x_1, \ldots, x_k, X_{k+1}, \ldots, X_n)] = \int_{\mathbb{R}^{n-k}} \phi(x_1, \ldots, x_n) dF_{k+1}(dx_{k+1}) \ldots dF_n(dx_n).
\]

Assume that \( E[\phi(X_1, \ldots, X_n)] < \infty \). Then we have that

\[
E[\phi(X_1, \ldots, X_n)](X_1, \ldots, X_k) = g(X_1, \ldots, X_k).
\]

Proof. We must first assess whether \( g(X_1, \ldots, X_k) \in \sigma(X_1, \ldots, X_k) \). As \( g \) is measurable (by Fubini’s theorem), then for every \( A \in \mathcal{B}(\mathbb{R}) \) we have that

\[
\{ \omega \in \Omega : g(X_1(\omega), \ldots, X_k(\omega)) \in A \} = \{ \omega \in \Omega : (X_1(\omega), \ldots, X_k(\omega)) \in g^{-1}(A) \} \in \sigma(X_1, \ldots, X_k),
\]

by definition. Now it remains to show that for every \( A \in \sigma(X_1, \ldots, X_k) \) that

\[
\int_A \phi(X_1, \ldots, X_n) dP = \int_A g(X_1, \ldots, X_k) dP.
\]
Now, as $A \in \sigma(X_1, \ldots, X_k)$ then $A = \{\omega \in \Omega : (X_1(\omega), \ldots, X_k(\omega)) \in C\}$ for some $C \in \mathcal{B}(\mathbb{R}^k)$, so we have

\[
\int_A \phi(X_1, \ldots, X_n) \, dP = \int_\Omega \phi(X_1, \ldots, X_n) 1_C(X_1, \ldots, X_k) \, dP \\
= \int_{\mathbb{R}^n} \phi(x_1, \ldots, x_n) 1_C(x_1, \ldots, x_k) F_1(dx_1) \cdots F_n(dx_n) \\
= \int_{\mathbb{R}^k} 1_C(x_1, \ldots, x_k) F_1(dx_1) \cdots F_k(dx_k) \int_{\mathbb{R}^{n-k}} \phi(x_1, \ldots, x_n) F_{k+1}(dx_{k+1}) \cdots F_n(dx_n) \\
= \int_{\mathbb{R}^k} g(x_1, \ldots, x_k) 1_C(x_1, \ldots, x_k) F_1(dx_1) \cdots F_k(dx_k) \\
= \int_A g(X_1, \ldots, X_k) \, dP.
\]

This is because of independence, independence and Fubini’s theorem. Note that if $A = \mathbb{R}^k$ then we get the useful result, 

\[
E[\phi(X_1, \ldots, X_n)] = \int_{\mathbb{R}^k} g(X_1, \ldots, X_k) F_1(dx_1) \cdots F_k(dx_k).
\]

\[\square\]

A NOTE ON DENSITIES—If $\mu$ and $\nu$ are $\sigma$-finite measures on a measurable space $(X, \mathcal{B})$. If $\nu$ is absolutely continuous with respect to $\mu$ then there exists a function $f : X \to [0, \infty]$, measurable with respect to $\mathcal{B}$, such that for $A \in \mathcal{B}$ such that

\[
\nu(A) = \int_A f(x) \, d\mu(x).
\]

Scheffé’s theorem. Suppose that probability measures $P_n$ and $P$ have densities $f_n$ and $f$ with respect to a measure $\mu$ on measurable space $(X, \mathcal{B})$ where $X$ a metric space. Then we have that

\[
d_{TV}(P_n, P) = \sup_{A \in \mathcal{B}} |P_n(A) - P(A)| = \frac{1}{2} \int_X |f_n(x) - f(x)| \, d\mu(dx) \to 0,
\]

as $n \to \infty$ if $f_n(x) \to f(x)$ a.s. with respect to $\mu$.

Proof. We know that

\[
P_n(X) - P(X) = \int_X f_n(x) - f(x) \, d\mu(dx) = 0,
\]

and hence for any $A \in \mathcal{B}$ we have that

\[
0 = \int_A f_n(x) - f(x) \, d\mu(dx) + \int_{A^c} f_n(x) - f(x) \, d\mu(dx) \\
\Leftrightarrow \int_A f_n(x) - f(x) \, d\mu(dx) = -\int_{A^c} f_n(x) - f(x) \, d\mu(dx) \\
\Rightarrow |\int_A f_n(x) - f(x) \, d\mu(dx)| = |\int_{A^c} f_n(x) - f(x) \, d\mu(dx)|.
\]

Therefore, we have that

\[
2|P_n(A) - P(A)| = 2|\int_A f_n(x) - f(x) \, d\mu(dx)| \\
= |\int_A f_n(x) - f(x) \, d\mu(dx)| + |\int_{A^c} f_n(x) - f(x) \, d\mu(dx)| \\
\leq \int_A |f_n(x) - f(x)| \, d\mu(dx) + \int_{A^c} |f_n(x) - f(x)| \, d\mu(dx) \\
= \int_X |f_n(x) - f(x)| \, d\mu(dx).
\]
Taking supremums, we get that
\[
\sup_{A \in \mathcal{B}} |P_n(A) - P(A)| \leq \frac{1}{2} \int_X |f_n(x) - f(x)| \mu(dx).
\]

Now, as \(f, f_n\) measurable we have that \(f_n - f\) is as well. Therefore, if \(B := \{ x \in X : f_n - f \geq 0 \} \in \mathcal{B}\), we get that
\[
2 \sup_{A \in \mathcal{B}} |P_n(A) - P(A) \geq |P_n(B) - P(B)|
\]
\[
= | \int_B f_n(x) - f(x) \mu(dx) | + | \int_{B^c} f_n(x) - f(x) \mu(dx) |
\]
\[
= \int_B |f_n(x) - f(x)| \mu(dx) + \int_{B^c} |f_n(x) - f(x)| \mu(dx)
\]
\[
= \int_X |f_n(x) - f(x)| \mu(dx).
\]

Thus, we have proved equality. Now it simply remains to show our convergence result. Now suppose that \(f_n(x) \to f(x) \) \(-\text{a.s.}\). Then we have that \(|f_n – f| \to 0 \) \(-\text{a.s.}\). Hence, \((f_n – f)^+ := \max(f_n – f, 0) \to 0 \) \(-\text{a.s.}\). Using the fact that for any real-valued function \(g\), we have that \(g = g^+ - g^-\) and \(|g| = g^+ + g^-\) then we get that
\[
\int_X (f(x) - f_n(x))^+ dx = \int_X (f(x) - f_n(x))^- dx,
\]
and therefore
\[
\int_X |f(x) - f_n(x)| \mu(dx) = \int_X (f(x) - f_n(x))^+ \mu(dx) + \int_X (f(x) - f_n(x))^- \mu(dx)
\]
\[
= 2 \int_X (f(x) - f_n(x))^+ \mu(dx).
\]

Now by \((f – f_n)^+ \leq f \in L^1\) and \((f – f_n)^+ \to 0 \) \(-\text{a.s.}\) the dominated convergence theorem gives us that
\[
\int_X |f_n(x) - f(x)| \mu(dx) \to 0,
\]
as \(n \to \infty\).

\[\square\]

**Inverse distribution function properties.** Let \(\mu\) be a measure on the Borel subspace \((A, \mathcal{B}(A))\) of \(\mathbb{R}\). Then \(F(x) := \mu((-\infty, x]\cap A)\) is nondecreasing and right-continuous. Furthermore, \(A_y := \{ x \in A : F(x) \geq y \}\) is closed and \(F^- (y) := \inf_{A_y}\) is such that \(F^- (y) \leq x\) if and only if \(y \leq F(x)\).

**Proof:** We note that if \(x_0 \leq x\), then \(\mu((-\infty, x_0]\cap A) \leq \mu((-\infty, x]\cap A)\) by monotonicity of \(\mu\). Now let \(x_n \downarrow x\). As \(\bigcap_n (-\infty, x_n] = (-\infty, x]\) then by continuity from above, we have that \(F(x_n) = \mu((-\infty, x_n]\cap A) \downarrow \mu((-\infty, x]\cap A) = F(x)\). Thus \(F\) is right-continuous.

Now, let us take \(x \in \bar{A}_y\). As \(A\) a metric subspace of \(\mathbb{R}\) in the restricted metric \(d : A \times A \to \mathbb{R}_+\) defined by \(d(a, b) = |a – b|\), then we have the existence of a sequence \(x_n \in \bar{A}_y\) such that \(x_n \to x\). Let us choose \(\epsilon > 0\) and note that by right-continuity there exists a \(\delta > 0\) such that \(F(x + \delta) – F(x) < \epsilon\). Furthermore, there exists an \(N\) such that for \(n \geq N\) we have \(|x_n - x| < \delta\). Suppose that \(x_n \leq x\). Then we have that
\[
F(x) + \epsilon \geq F(x) \geq F(x_n) \geq y.
\]

Now suppose that \(x_n \geq x\). By construction, we have \(x_n < x + \delta\), so that \(F(x_n) \leq F(x + \delta)\) and \(F(x) - F(x_n) < \epsilon\). Therefore,
\[
y \leq F(x_n) \leq F(x) + \epsilon.
\]
Hence, \(F(x) + \epsilon \geq y\) for every \(\epsilon > 0\). However, as \(\epsilon\) arbitrary we have that \(F(x) \geq y\) so \(x \in A_y\). Hence \(A_y\) closed.
As $A_y$ closed then we have $F^-(y) = \inf A_y \in A_y$, and so $y \leq F(F^-(y))$. Suppose that $y \leq F(x)$. Thus, $x \in A_y$ and so $F^-(y) \leq x$. There are at least two ways to show the converse. In the first, take $F^-(y) \leq x$. Then we have $y \leq F(F^-(y)) \leq F(x)$ by $F$ nondecreasing.

In the second, we use the right-continuity of $F$ and prove the contrapositive of the converse. Let us assume that $y > F(x)$. By right-continuity, there exists a $\delta > 0$ such that $y > F(x+\delta)$ and so $x+\delta \notin A_y$. This implies that $F^-(y) \geq x + \delta > x$. □

**Convergence in probability and equality in distribution.** Let us define random variables $X_n, X, Y_n,$ and $Y$ on $(\Omega, F, P)$ such that $X_n \overset{p}{\to} X$, $X \overset{d}{=} Y$ and $X_n \overset{d}{=} Y_n$ for all $n$. Then

$$Y_n \overset{d}{\to} Y.$$ 

Proof. Let $\epsilon > 0$, then

$$P(X_n \leq x) = P(X_n \leq x, |X_n - X| < \epsilon) + P(X_n \leq x, |X_n - X| \geq \epsilon)$$

$$\leq P(X < x + \epsilon) + P(|X_n - X| \geq \epsilon),$$

thus for every $\epsilon > 0$ we have

$$\limsup_{n \to \infty} P(X_n \leq x) \leq P(X < x + \epsilon).$$

Hence,

$$\limsup_{n \to \infty} P(X_n \leq x) \leq P(X \leq x),$$

and equality in distribution supplies us with

$$\limsup_{n \to \infty} P(Y_n \leq x) \leq P(Y \leq x).$$

Now, we consider a similar inequality. Namely, if we let $\epsilon > 0$ then

$$P(X \leq x - \epsilon) \leq P(X_n \leq x) + P(|X_n - X| \geq \epsilon),$$

which is equivalent to

$$P(X \leq x - \epsilon) + \liminf_{n \to \infty} P(|X_n - X| \geq \epsilon) \leq \liminf_{n \to \infty} P(X_n \leq x),$$

so by applying the equivalence in distributions again we have

$$P(Y \leq x) \leq \liminf_{n \to \infty} P(Y_n \leq x) \leq \limsup_{n \to \infty} P(Y_n \leq x) \leq P(Y \leq x).$$

If $x$ a continuity point of $F(x) = P(Y \leq x)$, then $F(x) - F(x-) = 0$ thus $P(Y < x) = P(Y \leq x)$ and hence $Y_n \overset{d}{\to} Y$ by definition of weak convergence. (Recall, that the system of rays of the form $(-\infty, x]$ are a convergence-determining class on the Borel $\sigma$-algebra on $\mathbb{R}$, and that $(P \circ Y^{-1})(\partial(-\infty, x]) = P(Y \leq x) - P(Y < x)$, corresponding to the more general definition of weak convergence on metric spaces, evinced in [5]). □

**Stochastic domination and nonnegative random variables.** Let $X, Y$ be nonnegative random variables defined on $(\Omega, F, P)$ such that for all $t \geq 0$ we have $P(Y \geq t) \geq P(X \geq t)$. Then if $f : \mathbb{R} \to [0, \infty)$ is a nondecreasing continuously differentiable (i.e. $C^1$) function such that $f(0) = 0$ we have that $E[f(Y)] \geq E[f(X)]$.

Proof. It is sufficient to show that for any nonnegative random variable $X$ that

$$E[f(X)] = \int_0^\infty f'(t)P(X \geq t) \ dt.$$
To show this, we first see that

\[
E[f(X)] = \int_{\Omega} f(X(\omega)) \, dP
\]

\[
= \int_{\Omega} \int_0^{X(\omega)} f'(t) \, dt \, dP
\]

\[
= \int_{\Omega} \int_0^\infty f'(t) 1_{\{X(\omega) \geq t\}} \, dt \, dP
\]

\[
= \int_0^\infty \int_{\Omega} f'(t) 1_{\{X(\omega) \geq t\}} \, dP \, dt
\]

\[
= \int_0^\infty f'(t) P(X \geq t) \, dt,
\]

and the desired conclusion follows from the monotonicity of the Lebesgue integral. For \( X \) a non-negative integer-valued random variable, this becomes

\[
E[f(X)] = \sum_{n=1}^{\infty} [f(n) - f(n - 1)] P(X \geq n).
\]

References


