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 $1 + x \geq 0$,

\[
(1 + x)^{n+1} = (1 + x)(1 + x)^n \geq (1 + x)(1 + nx) = 1 + x + nx + nx^2 = 1 + (n + 1)x + nx^2 \geq 1 + (n + 1)x.
\]

\[
\Box
\]

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

\[
(1 + \frac{x}{n})^n = \sum_{k=0}^{n} \binom{n}{k} \frac{x^k}{n^k} = \sum_{k=0}^{n} \frac{x^k}{k!} \left(\frac{n(n-1) \cdots (n-k+1)}{n^k}\right)
\]

\[
\leq \sum_{k=0}^{n} \frac{x^k}{k!} \leq \sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x.
\]

Furthermore, for $n \geq m$ we have by $(\frac{n}{k}) \frac{x^k}{n^k} \geq 0$ that

\[
(1 + \frac{x}{n})^n \geq \sum_{k=0}^{m} \binom{n}{k} \frac{x^k}{n^k} =: s_{n,m}.
\]

Now, as

\[
\left(\frac{n(n-1) \cdots (n-k+1)}{n^k}\right) = \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \to 1,
\]

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
\[
\lim_{n \to \infty} \left( 1 - \frac{x}{n} \right)^n = \lim_{n \to \infty} \left( \frac{1 - \frac{x^2}{n^2}}{1 + \frac{x}{n}} \right)^n = e^{-x}.
\]

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

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

**Proof.** First note that for $x > 0$ we have that
\[
\frac{1}{(x+1)^2} \leq \frac{1}{x(x+1)},
\]
thus we have
\[
\int_x^\infty \frac{1}{(t+1)^2} \, dt \leq \int_x^\infty \frac{1}{t(t+1)} \, dt = \lim_{y \to \infty} \int_x^y \frac{1}{t} - \frac{1}{t+1} \, dt.
\]
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 $e \leq (1 + 1/x)^{x+1}$ for any $x > 0$ – in particular the natural numbers. This is equivalent to showing that $(x+1) \log(1 + 1/x) \geq 1$ or rather that $\log(1 + 1/x) \geq 1/(1 + x)$, which is exactly the result from above.

**A technical result for the DeMoivre-Laplace theorem.** Let $0 \leq p \leq 1$. Furthermore, suppose that for $k_n(x)$ is a sequence of integers such that $|k_n(x)| \leq n$ 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 $x \mapsto e^x$ a continuous function. If we recognize that
\[
\frac{np}{k_n} = 1 - \frac{k_n - np}{k_n} \quad \text{and} \quad \frac{n(1-p)}{n-k_n} = 1 + \frac{k_n - np}{n-k_n},
\]
then we will use the fact that the Maclaurin series for $\log(1 + x)$ is

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

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 $x \to 1$ is continuous and $k_n/np \to 1$. The other case follows similarly. To demonstrate $k_n/np \to 1$ just take

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

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

$$\log(1 + x) = \int_{0}^{x} \frac{1}{1+y} \, dy \leq \int_{0}^{x} 1 \, dy = x,$$

which implies our result – because for $x \leq 0$, we have $1 + x \leq 0 \leq e^x$. 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 $(1 + x)^{1+x} \leq e^{x+x^2/2}$, so that we get $e^{-x^2/2} (1 + x)^{1+x} \leq e^x$ or $e^{-x^2/2} \leq e^x (1 + x)^{-(1+x)}$.
which implies that if \( U \) is an open cover of \( X \) then we have equality. If \( m = 1 \) and \( k \geq 2 \) then
\[
\binom{n + m}{k + 1} - \binom{n}{k + 1} = \frac{(n + 1)(n + 1 - k) - n(n - 1) \cdots (n - k)}{(k + 1)!} \leq \frac{n(n - 1) \cdots (n - k + 1)(n + 1 - (n - k))}{(k + 1)!} = \binom{n}{k} \leq \frac{n^k}{k!}.
\]
Now if we assume the desired inequality holds for \( m \), then
\[
\binom{n + m + 1}{k + 1} - \binom{n}{k + 1} = \left( \binom{n + m}{k + 1} + \binom{n + m}{k} \right) - \binom{n}{k + 1} \leq \frac{n + m}{k} + \frac{m(n + m)^k}{k!} \leq \frac{(m + 1)(n + m + 1)^k}{k!},
\]
hence proved.

Unaccompanied lemmas

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
\[
C \subset \bigcup_{i=1}^{n} U_{\alpha_i} = f^{-1}(\bigcup_{i=1}^{n} V_{\alpha_i}),
\]
which implies that if \( x \in C \), then \( f(x) \in \bigcup_{i=1}^{n} V_{\alpha_i} \). Then as \( y = f(x) \in f(C) \) for some \( x \in C \) by definition this entails that \( f(C) \subset \bigcup_{i=1}^{n} V_{\alpha_i} \) and hence \( f(C) \) compact.

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

Proof. We first mention that if \( s_n \in \mathbb{R} \) and \( s_* = \liminf s_n \) then for every \( s < s_* \) there exists an \( N_y \) such that \( s_n > y \) for \( n \geq N_y \). Now, suppose that \( \liminf_{n \to \infty} 1_{A_n}(x) = 1 \). 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, \( x \in \liminf A_n \) so that \( 1 \liminf_{n \to \infty} A_n(x) = 1 \).
Now suppose that $1_{\liminf_{n \to \infty} A_n}(x) = 1$, then by definition of limit infimum for sets, there exists some $N$ such that for all $n \geq N$ 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 $\liminf_{n \to \infty} 1_{A_n}(x) = 1$. Thus

$$1_{\liminf_{n \to \infty} A_n}(x) = 1 \iff \liminf_{n \to \infty} 1_{A_n}(x) = 1,$$

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

$$d(f_n(x_n), f(x)) \leq d(f(x), f(x_n)) + d(f(x_n), f_n(x_n)) < \epsilon/2 + \epsilon/2 = \epsilon.$$

**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 $s : \mathcal{P}(X) \to \mathcal{P}(\mathbb{R})$ 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 $\mu^*(E_{n}) + \epsilon 2^{-n} > \sum_{m=1}^{\infty} \mu_0(E_{n,m})$, 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.

**Uniform continuity and modulus.** Let $f : X \to Y$ 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 $\omega_{\delta}(f) \to 0$ as $\delta \to 0$.

**Proof.** Suppose that $f$ is uniformly continuous. Let us take $\delta_n \to 0$ and suppose that $\omega_{\delta_n}(f) \not\to 0$. 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. \( \Box \)

**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 converge 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. \( \Box \)

**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. \( \Box \)
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 $f_i$ 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. □

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 $m(T(A)) = |\det(T)| m(A), \quad$
Proof. Let because identical sequences must have the same limit – or rather \( \lim A \). Suppose that \( A \) is 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
\[
\int_{\mathbb{R}^d} f(T^{-1}(x)) \, dx = \int_{\mathbb{R}^d} \sum_{i=1}^n c_i 1_{A_i}(T^{-1}(x)) \, dx = \sum_{i=1}^n c_i \int_{\mathbb{R}^d} 1_{T(A_i)}(x) \, dx
\]
\[
= |\det(T)| \sum_{i=1}^n c_i m(A_i) = |\det(T)| \int_{\mathbb{R}^d} \sum_{i=1}^n c_i 1_{A_i}(x) \, dx = |\det(T)| \int_{\mathbb{R}^d} f(x) \, dx.
\]
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.

The probabilistic touch

Separating class theorem. Let probability measures \( P \) and \( Q \) be defined on the measurable space \((\Omega, F)\). 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 \( F = \sigma(\mathcal{C}) \) for some collection \( \mathcal{C} \) of sets, then \( P(A) = Q(A) \) for all \( A \in F \) 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 \( A_1, A_2, \ldots \in \mathcal{L} \). 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 if \( \sigma(\mathcal{C}) \) minimal, then \( \sigma(\mathcal{C}) \subset \sigma(\mathcal{P}) \) and so \( P \) and \( Q \) coincide on \( F \).

Countable totality. Let us take the probability space \((\Omega, F, P)\) and let \( \{A_n\} \) be an uncountable family of disjoint subsets of \( \Omega \) measurable with respect to \( F \). Then at most only countably many sets in \( \{A_n\} \) 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 \( P(A_\alpha) > 0 \). 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.
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 $A, B \in \mathcal{B}(\mathbb{R})$

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

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

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

□

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

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

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

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

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

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 $X_1$ 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

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

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

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

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

$$\int_A g(X_1, \ldots, X_k) \, dP = \int_A \phi(X_1, \ldots, X_n) \, dP.$$
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,

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

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 \( f : X \rightarrow [0, \infty] \), 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

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

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

**Proof.** We know that

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

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

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

\[
\Leftrightarrow \int_A f_n(x) - f(x) \, d\mu(x) = -\int_{A^c} f_n(x) - f(x) \, d\mu(x)
\]

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

Therefore, we have that

\[
2|P_n(A) - P(A)| = 2|\int_A f_n(x) - f(x) \, d\mu(x)|
\]

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

\[
\leq \int_A |f_n(x) - f(x)| \, d\mu(x) + \int_{A^c} |f_n(x) - f(x)| \, d\mu(x)
\]

\[
= \int_X |f_n(x) - f(x)| \, d\mu(x).
\]
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
\[ 2 \sup_{A \in \mathcal{B}} |P_n(A) - P(A)| \geq 2 |P_n(B) - P(B)| \]
\[ = | \int_B f_n(x) - f(x) \mu(dx) | + | \int_{B^c} f_n(x) - f(x) \mu(dx) | \]
\[ = \int_B |f_n(x) - f(x)| \mu(dx) + \int_{B^c} |f_n(x) - f(x)| \mu(dx) \]
\[ = \int_X |f_n(x) - f(x)| \mu(dx). \]
Thus, we have proved equality. Now it simply remains to show our convergence result. Now suppose that \( f_n(x) \to f(x) \) \( \mu \text{-a.s.} \). Then we have that \( f - f_n \to 0 \) \( \mu \text{-a.s.} \). Hence, \( (f - f_n)^+ := \max(f_n - f, 0) \to 0 \) \( \mu \text{-a.s.} \). 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 \( (f - f_n)^+ \to 0 \) \( \mu \text{-a.s.} \), the dominated convergence theorem gives us that
\[ \int_X |f_n(x) - f(x)| \mu(dx) \to 0, \]
as \( n \to \infty. \)

**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 \( \cap_n (-\infty, x_n] = (-\infty, x] \) 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 \( x \in A_y \). 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 \( x_n \in A_y \) 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 \( F(x_n) - F(x) < \epsilon \). 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 $F^r(y) = \inf A_y \in A_y$, and so $y \leq F(F^r(y))$. Suppose that $y \leq F(x)$. Thus, $x \in A_y$ and so $F^r(y) \leq x$. There are at least two ways to show the converse. In the first, take $F^r(y) \leq x$. Then we have $y \leq F(F^r(y)) \leq F(x)$ 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 $F^r(y) \geq x + \delta > x$.

### Convergence in probability and equality in distribution

Let us define random variables $X_n, X, Y_n$, and $Y$ on $(\Omega, \mathcal{F}, P)$ 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

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

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 $Y_n \overset{d}{=} Y$ 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 $(\Omega, \mathcal{F}, P)$ 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
\[
E[f(X)] = \int_{\Omega} f(X(\omega)) \, dP \\
= \int_{\Omega} \int_{0}^{X(\omega)} f'(t) \, dt \, dP \\
= \int_{0}^{\infty} \int_{\Omega} f'(t) \mathbf{1}_{\{X(\omega) \geq t\}} \, dt \, dP \\
= \int_{0}^{\infty} \int_{\Omega} f'(t) \mathbf{1}_{\{X(\omega) \geq t\}} \, dP \, dt \\
= \int_{0}^{\infty} f'(t) P(X \geq t) \, dt,
\]
and the desired conclusion follows from the monotonicity of the Lebesgue integral. For $X$ a non-negative integer-valued random variable, this becomes
\[
E[f(X)] = \sum_{n=1}^{\infty} [f(n) - f(n-1)] P(X \geq n).
\]

\[\square\]

References