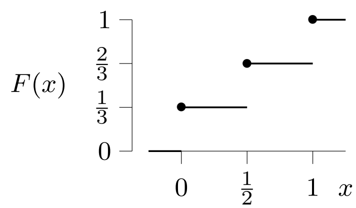
TOPIC Distribution functions and their inverses. This section develops properties of probability distribution functions and their inverses. Two main topics are the so-called probability integral transformation and inverse probability transformation.

Distribution Functions. Let X be a real-valued random variable defined on a sample space Ω. The ***distribution function*** (***df***) of X is the function $F \equiv F_X$ from $\mathbb{R} := (-\infty, \infty)$ to $[0,1]$ defined by

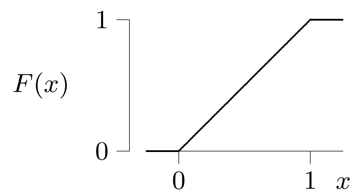
$$F(x) := P[\,\omega \in \Omega : X(\omega) \le x\,] \equiv P[\,X \le x\,].$$

Here are a couple of examples which motivate Theorem 1 below. The symbol "$\sim$" is to be read as "distributed as":

df of $X \sim$ Uniform on $\{0, 1/2, 1\}$



df of $X \sim$ Uniform on $[0,1]$



Theorem 1 (Properties of F). *For any random variable X, the distribution function F of X has these properties:*

DF1 *F is nondecreasing;*

DF2 *F is right-continuous;*

DF3 $\lim_{x\to-\infty} F(x) = 0$ *and* $\lim_{x\to\infty} F(x) = 1$.

Moreover for each $x \in \mathbb{R}$,

DF4 $F(x-) := \lim_{w\uparrow x,\, w<x} F(w) = P[\,X < x\,]$, *and*

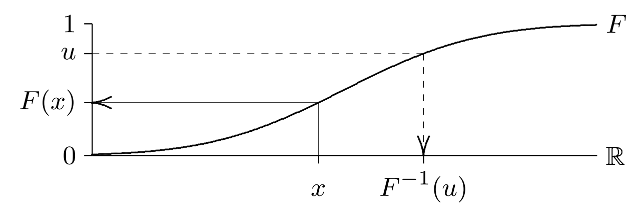
DF5 *jump of F at $x := F(x) - F(x-) = P[\,X = x\,]$.*

Proof I will prove DF1 and DF2, and leave the rest to you as Exercise 1.

• DF1 $x \le y \implies F(x) \le F(y)$: Indeed, suppose $x \le y$. Then the event $A := \{X \le x\}$ is contained in the event $B := \{X \le y\}$, so $F(x) = P[A] \le P[B] \le F(y)$.

• DF2 $x_n \downarrow x \implies F(x_n) \downarrow F(x)$. Indeed, suppose $x_1, x_2, \ldots$ is an infinite sequence of real numbers that decrease down to x. Then the events $A_n := \{X \le x_n\}$ "shrink down" to the event $A := \{X \le x\}$. By a property of probability measures (see (15), below), $F(x_n) = P[A_n]$ decreases down to $P[A] = F(x)$. ∎

A special case. This section discusses inverse dfs and the probability integral and inverse probability transformations under the simplifying assumption that the distribution function F of X is continuous and strictly increasing, as illustrated below:



For $u \in (0,1)$, let $F^{-1}(u)$ be defined as in the picture, i.e., $F^{-1}(u)$ is the unique number ξ such that $F(\xi) = u$. By a result in analysis, $F^{-1}(u)$ is continuous and strictly increasing in u.

Consider now the random variable $F(X)$, whose value at a sample point ω is $F\big(X(\omega)\big) = P[\,\omega' : X(\omega') \le X(\omega)\,]$, the probability of observing a new value for X no greater than the value $X(\omega)$ at hand. For $0 < u < 1$, we have $F(X) \le u \iff X = F^{-1}\big(F(X)\big) \le F^{-1}(u)$, so

$$P[\,F(X) \le u\,] = P[\,X \le F^{-1}(u)\,] = F\big(F^{-1}(u)\big) = u. \tag{1}$$

This implies that $F(X)$ is uniformly distributed over $(0,1)$: in symbols, $F(X) \sim U$ for $U \sim$ Uniform(0,1).

(1): $F(X) \sim U \sim \text{Uniform}(0{,}1)$.

---

(1) has implications for statistics, in the context of hypothesis testing. Think of X as a test statistic for the simple hypothesis H that the data are distributed according to P, with the alternative A such that you ought to reject H when X is too far to the left. For an observed value x of X, $F(x) = P[X \le x]$ is the chance of getting a result as extreme, or more so, than the one at hand. In statistics, this quantity is called the ***p-value***. Small p-values argue against H (and thus in favor of A); in decision theory you reject H (and accept A) if the p-value is sufficiently small, say ≤ 0.05. (1) says that when H is in fact true, repeated tests will produce p-values that are uniformly distributed over $(0,1)$; by chance alone, you'll get a p-value less than 0.05 (and mistakenly reject H) about 1 time in 20.

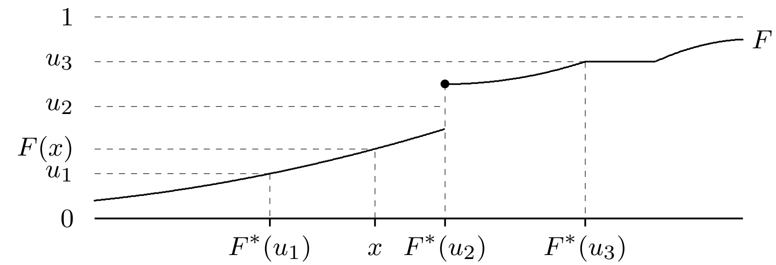
Now let U be a random variable uniformly distributed over $(0,1)$ and consider the random variable $F^{-1}(U)$. For $x \in \mathbb{R}$, $F^{-1}(U) \le x \iff U = F\big(F^{-1}(U)\big) \le F(x)$, so

$$P[\,F^{-1}(U) \le x\,] = P[\,U \le F(x)\,] = F(x) = P[\,X \le x\,]. \tag{2}$$

Since this is true for all x, $F^{-1}(U)$ and X have the same distribution: $F^{-1}(U) \sim X$. This result has implications for random number generation. Namely, if you can somehow generate a uniform variable U, then $F^{-1}(U)$ will be distributed like X. In principle this method can always be used to simulate X, but it is efficient only when F^{-1} is easy to compute. When that's not the case, one can often get an efficient algorithm by using some result from distribution theory.

The transformation from X to $F(X)$ is called the ***probability integral transformation (PIT)***, whereas the transformation from U to $F^{-1}(U)$ is called the ***inverse probability transformation (IPT)***. In what follows we are going to study the PIT and IPT in the general case, where F may be discontinuous and not strictly increasing.

Inverse distribution functions. Let F be an arbitrary probability distribution function. Since the graph of F can have jumps and flat spots, there is no true inverse to F in the usual sense. One can define a kind of inverse F^* to F, as follows. Refer to the figure below:



As a first attempt, try taking $F^*(u)$ to be that x such that $u = F(x)$. This definition works for $u = u_1$, but it doesn't work for $u = u_3$, for which there is a whole range of x's such that $u = F(x)$. As a second attempt, try taking $F^*(u)$ to be the smallest x such that $u = F(x)$. This works for $u = u_3$ and u_1, but it doesn't work for $u = u_2$, since there are no x's such that $u_2 = F(x)$. As a third attempt, try taking $F^*(u)$ to be the smallest x such that $u \le F(x)$. This works for $u = u_2$ and u_3 and u_1. We make this the general definition: more precisely we take

$$F^*(u) := \inf\{\, x \in \mathbb{R} : u \le F(x)\,\} \tag{3}$$

for $u \in (0,1)$; here "inf" stands for "infimum", or "greatest lower bound". To better understand (3) fix $u \in (0,1)$ and consider the set

$$I := \{x \in \mathbb{R} : u \le F(x)\,\}.$$

Note that: I is nonempty, because $u < 1$ and $F(x) \to 1$ as $x \to \infty$; I is an interval extending out to $+\infty$, because F is nondecreasing; I has a finite left-endpoint, say ξ, because $u > 0$ and $F(x) \to 0$ as $x \to -\infty$; and $\xi \in I$, because F is right-continuous. The last claim follows from

$$\begin{aligned} x_n := \xi + 1/n \in I \ \text{ for all integers } n \ge 1 &\Longrightarrow u \le F(x_n) \ \text{ for all } n \\ &\Longrightarrow u \le \lim\nolimits_n F(x_n) = F(\xi) \Longrightarrow \xi \in I. \end{aligned}$$

To summarize, $I = \{\, x \in \mathbb{R} : u \le F(x) \,\}$ is a left-closed right-semi-infinite interval and $F^*(u) = \xi$ is its finite left endpoint. This gives

Theorem 2. *Let F be a probability distribution function and let $F^*\colon (0,1) \to \mathbb{R}$ be defined by $F^*(u) = \inf\{x \in \mathbb{R} : u \le F(x)\,\}$. The infimum here is attained:*

$$F^*(u) \text{ is in fact the smallest } x \in \mathbb{R} \text{ such that } u \le F(x). \tag{4}$$

Moreover, for any $u \in (0,1)$ and $x \in \mathbb{R}$, one has

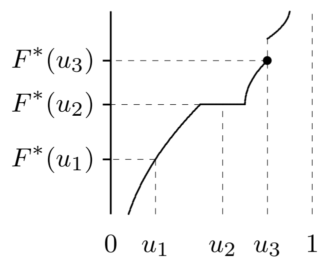
$$u \le F(x) \iff F^*(u) \le x. \tag{5}$$

Relation (5) is called the ***switching formula*** (SF). (5) has an obvious counterpart:

$$u > F(x) \iff F^*(u) > x; \tag{6}$$

But watch out! If "$\le$" is changed to "$<$" throughout (5), or if "$>$" is changed to "$\ge$" throughout (6), the resulting assertions may not hold: see Exercises 2 and 8. In these notes, when invoking (5) and (6), I will always write the "u-thing" on the left and the "x-thing" on the right and only use the valid inequalities "$\le$" and "$>$".

The theorem below gives the main properties of F^*. To motivate them, here is the graph of $F^*(u)$ versus u for the F on the preceding page (the scales are different, though):



Note that this F^* is nondecreasing and left-continuous. As is the case for any nondecreasing function,

$$F^*(u+) := \lim_{v \downarrow u,\, v > u} F^*(v) \tag{7}$$

exists for each u.

(3): $F^*(u) = \inf\{\, x \in \mathbb{R} : u \le F(x) \,\}$. (SF): $u \le F(x) \iff F^*(u) \le x$

Theorem 3 (Properties of F^*). *Let F be a probability distribution function and let F^* be defined by* (3) *for $0 < u < 1$. Then*

- IDF1 *F^* is nondecreasing;*
- IDF2 *F^* is left-continuous;*
- IDF3 $\lim_{u \downarrow 0} F^*(u) = \inf\{\, x \in \mathbb{R} : F(x) > 0 \,\}$ *and* $\lim_{u \uparrow 1} F^*(u) = \sup\{\, x \in \mathbb{R} : F(x) < 1 \,\}$.
- IDF4 *for each $u \in (0,1)$ and $x \in \mathbb{R}$ with $0 < F(x) < 1$,*

$$F\big((F^*(u))-\big) \le u \le F\big(F^*(u)\big), \tag{8}$$

$$F^*\big(F(x)\big) \le x \le F^*\big((F(x))+\big). \tag{9}$$

Proof • IDF1 $0 < u \le v < 1 \implies F^*(u) \le F^*(v)$: This follows easily (show how!) from the definition of F^*. It also follows from the switching formula:

$$\begin{aligned} F^*(v) \le F^*(v) &\implies v \le F\big(F^*(v)\big) && \text{(by the SF)} \\ &\implies u \le F\big(F^*(v)\big) && \text{(since } u \le v\text{)} \\ &\implies F^*(u) \le F^*(v) && \text{(SF again).} \end{aligned}$$

• IDF2: $u_n \uparrow u \implies F(u_n) \uparrow F(u)$. The assumption is $u_n \le u_{n+1}$ for all n and $u = \lim_n u_n$. Since F^* is nondecreasing, we have

$$F^*(u_n) \le F^*(u_{n+1}) \le F^*(u)$$

for all n, so $L := \lim_n F^*(u_n)$ exists and satisfies $L \le F^*(u)$. To get the opposite inequality, consider an $x \in \mathbb{R}$ with $F^*(u) > x$. Then

$$\begin{aligned} & u > F(x) && \text{(by the SF)} \\ \implies & u_n > F(x) \text{ for all large } n && \text{(since } u_n \uparrow u\text{)} \\ \implies & F^*(u_n) > x \text{ for all large } n && \text{(SF again)} \\ \implies & L > x && \text{(since } L = \lim_n F^*(u_n)\text{).} \end{aligned}$$

Now let x tend up to $F^*(u)$ to conclude $L \ge F^*(u)$, as desired.

• IDF3: This result is not so important, so I'll leave it to you as Exercise 3.

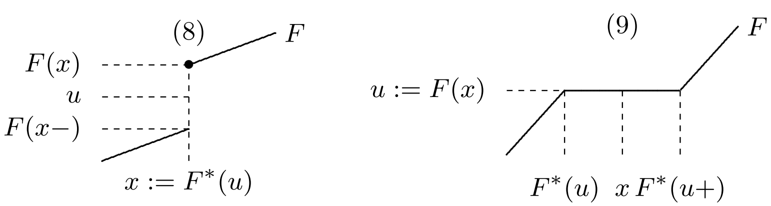
• (8) holds. The inequality on the right follows directly from the SF, as in the proof of IDF1. To get the inequality on the left, set $x = F^*(u)$; we need to show $u \ge F(x-)$. For this let $\xi < x$. Then

$$\begin{aligned} & F^*(u) > \xi && (\text{since } x = F^*(u)) \\ \Longrightarrow\ & u > F(\xi) && (\text{by the SF}). \end{aligned}$$

Letting ξ tend up to x shows that $u \ge F(x-)$, as desired.

• (9) holds. The argument for this is similar to that for (8); I'll leave it to Exercise 4. ∎

Relations (8) and (9) specify the extent to which F and F^* are inverses. Note that the inequalities in these relations can be strict, as in the following cases:



In view of IDF2 and IDF4, F^* is called the ***left-continuous inverse*** to F. IDF4 and the preceding examples yield this corollary:

Theorem 4. *Let F^* be the left-continuous inverse to the df F. Then*

$$F\big(F^*(u)\big) = u \text{ for all } u \in (0,1) \text{ iff } F \text{ is continuous,} \tag{10}$$

$$\left\{ \begin{matrix} F^*\big(F(x)\big) = x \text{ for all } x \in A := \{\, x \in \mathbb{R} : 0 < F(x) < 1 \,\} \\ \text{iff } F \text{ is strictly increasing over } A \end{matrix} \right\}. \tag{11}$$

(3): $F^*(u) = \inf\{\, x \in \mathbb{R} : u \le F(x) \,\}$. (SF): $u \le F(x) \iff F^*(u) \le x$

The inverse probability transformation. We saw earlier (see (1) and (2)) that if the df F of a random variable X is continuous and strictly increasing, then (i) $F(X)$ is uniformly distributed over $(0,1)$, and, conversely, (ii) if U is uniformly distributed over $(0,1)$, then $F^*(U) \sim X$. The following theorem says that (ii) without any conditions on F. (i) is not always true, but there are some things that can be said; we'll deal with that in the next subsection.

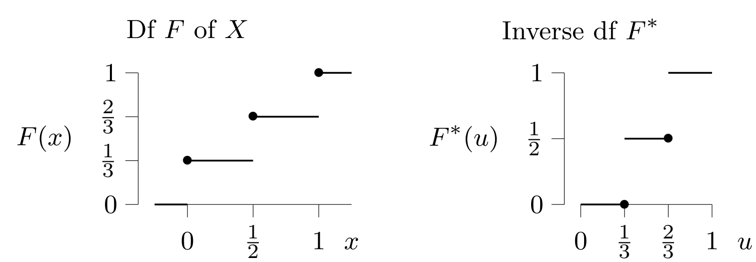
Theorem 5 (The IPT Theorem). *Let X be a random variable with df F and left-continuous inverse df F^*. If $U \sim (0,1)$, then $F^*(U) \sim X$.*

Proof For each $x \in \mathbb{R}$, we have $F^*(U) \le x \iff U \le F(x)$ by the switching formula. Thus

$$P[\, F^*(U) \le x \,] = P[\, U \le F(x) \,] = F(x) = P[\, X \le x \,].$$

Since this is true for all $x \in \mathbb{R}$, $F^*(U) \sim X$. ∎

Example 1. Suppose X takes the values 0, 1/2, and 1 with probability 1/3 each. The graphs of F and F^* are as follows:



It is clear that if $U \sim \text{Uniform}(0,1)$, then $F(U)$ takes the values 0, 1/2, and 1 with probability 1/3 each, just as X does. •

(8): $F\big(F^*(u)-\big) \le u \le F\big(F^*(u)\big)$. $\qquad U \sim \text{Uniform} \underset{\text{IPT}}{\Longrightarrow} F^*(U) \sim F$.

The probability integral transformation. The second half of the following theorem gives a necessary and sufficient condition for $F(X)$ to be uniformly distributed over $(0,1)$.

Theorem 6 (The PIT Theorem). *Let X be a random variable with df F. Then*

$$P[\,F(X) \le u\,] \le u \ \text{ for all } u \in (0,1). \tag{12}$$

Moreover

$$P[\,F(X) \le u\,] = u \ \text{ for all } u \in (0,1) \iff F \text{ is continuous.} \tag{13}$$

Proof Let $U \sim \text{Uniform}(0,1)$ and let F^* be the left-continuous inverse to F. By the IPT, $X \sim F^*(U)$, so

$$F(X) \sim F\big(F^*(U)\big).$$

• (12) *holds.* In general, $F\big(F^*(U)\big) \ge U$ by (8). Thus for all $u \in (0,1)$,

$$P[\,F(X) \le u\,] = P[\,F\big(F^*(U)\big) \le u\,] \le P[U \le u\,] = u.$$

• (13) *holds.* If F is continuous, then $U = F\big(F^*(U)\big)$ by (8), so

$$F(X) \sim F\big(F^*(U)\big) = U \sim \text{Uniform(0,1)}.$$

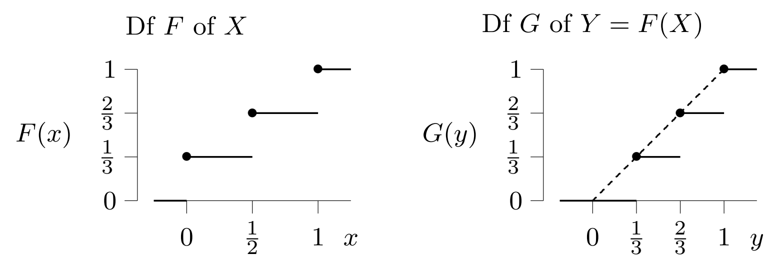
On the other hand, if F is not continuous, then there exists an $x \in \mathbb{R}$ such that

$$0 < F(x) - F(x-) = P[\,X = x\,] \le P[\,F(X) = F(x)\,].$$

But $P[\,U = F(x)\,] = 0$, so $F(X) \not\sim U$. ∎

(12): For $X \sim F$, $P[\,F(X) \le u\,] \le u$ for all $u \in (0,1)$.

Example 2. As in the preceding example, suppose X takes the values 0, 1/2, and 1 with probability 1/3 each. Then $Y := F(X)$ takes the values 1/3, 2/3, and 1 with probability 1/3 each. The graphs of the df F of X and the df G of Y are as follows:



The graph of G shows that $G(y) \le y$ for all $y \in (0,1)$, as (12) asserts.•

The proof of the PIT Theorem illustrates a useful technique — if you want to prove something about the distribution of a random variable X with df F, try representing X as $F^*(U)$ for a uniform random variable U.

Exercises. The following definitions and results are needed for Exercise 1. Suppose $(x_n)_{n=1}^{\infty}$ is an infinite sequence of real numbers and x is a real number. One writes

$$x_n \uparrow x \quad \text{to mean} \quad x_n \le x_{n+1} \text{ for all } n \text{ and } \lim_n x_n = x$$

and

$$x_n \downarrow x \quad \text{to mean} \quad x_n \ge x_{n+1} \text{ for all } n \text{ and } \lim_n x_n = x.$$

Similarly, if $(A_n)_{n=1}^{\infty}$ is an infinite sequence of events and A is an event, one writes

$$A_n \uparrow A \quad \text{to mean} \quad A_n \subset A_{n+1} \text{ for all } n \text{ and } A = \bigcup_{n=1}^{\infty} A_n,$$

and

$$A_n \downarrow A \quad \text{to mean} \quad A_n \supset A_{n+1} \text{ for all } n \text{ and } A = \bigcap_{n=1}^{\infty} A_n.$$

One of the properties of a probability measure P is that

$$A_n \uparrow A \Longrightarrow P[A_n] \uparrow P[A] \tag{14}$$

and

$$A_n \downarrow A \Longrightarrow P[A_n] \downarrow P[A]. \tag{15}$$

Properties (14) and (15) are called respectively ***continuity from below*** and ***continuity from above***.

Exercise 1. Complete the proof of Theorem 1, using (14) and (15) to verify properties DF3 and DF4. ⋄

Exercise 2. Let F^* be the left-continuous inverse to a df F. Show by examples that $F^*(u) < x$ does not imply $u < F(x)$ and that $u < F(x)$ does not imply $F^*(u) < x$. ⋄

Exercise 3. Prove IDF3. [Hint: put $L = \lim_{u \downarrow 0} F^*(u)$ and $\xi = \inf(A)$ for $A = \{ x \in \mathbb{R} : F(x) > 0 \}$. Note that L and ξ may be $-\infty$. Deduce $L \le \xi$ from the fact that $L \le x$ for each $x \in A$ (why?). Deduce $L \ge \xi$ from the fact that $F^*(u) \in A$ for each $u > 0$ (why?).] ⋄

Exercise 4. Prove (9). ⋄

Exercise 5. Let F be a df such that $0 < F(x) < 1$ for all $x \in \mathbb{R}$ and let F^* be the left-continuous inverse to F. Show that

$$F^*\big(F\big(F^*(u)\big)\big) = F^*(u) \text{ for all } u \in (0,1) \tag{16}$$

and

$$F\big(F^*\big(F(x)\big)\big) = F(x) \text{ for all } x \in \mathbb{R} \tag{17}$$ ⋄

Exercise 6. Prove Theorem 4. ⋄

Exercise 7. Inequality (12) has an important implication for p-values. What is that? ⋄

Exercise 8. Let X be a random variable with df F and left-continuous inverse df F^*. (a) Show that for $x \in \mathbb{R}$ and $0 < u < 1$, one has

$$u \ge F(x-) \iff F^*(u+) \ge x, \tag{18}$$

with $F(x-)$ defined as in DF4 and $F^*(u+)$ defined by (7). (b) Use part (a) to show that for $0 < u < 1$,

$$F^*(u+) = \sup\{ x : u \ge F(x-) \}. \tag{19}$$

[Hint for (a): use the switching formula $v > F(w) \iff F^*(v) > w$, noting for example that $u \ge F(x-) \iff u \ge F(w)$ for all $w < x$.] ⋄

Unfornately the jumps of F and F^* complicate what would otherwise be a simple theory. Forturnately, though, F and F^* don't have too many jumps — according to the following exercise, there are at most countably many of them.

Exercise 9. Let B be a subinterval of $\mathbb{R}$. (B doesn't have to be a proper subinterval; the case $B = \mathbb{R}$ is allowed.) Let f be a nondecreasing mapping from B into $\mathbb{R}$. (For example, f might be a df, defined on $B = \mathbb{R}$, or an inverse df, defined on $B = (0,1)$.) Put

$$\mathcal{D}_f := \{x \in B : f \text{ is discontinuous at } x\}, \tag{20}$$

$$\mathcal{C}_f := \mathcal{D}_f^c = \{x \in B : f \text{ is continuous at } x\}. \tag{21}$$

(a) Show that $\mathcal{D}_f$ is countable. (b) Use part (a) to show that $\mathcal{C}_f$ is dense in B. [Hint for (a): First show that for any closed bounded subinterval A of B and any number $\epsilon > 0$, there are at most finitely many points $x \in A$ such that the jump $f(x+) - f(x-)$ of f at x exceeds ϵ.] $\diamond$

The following exercise plays an important role in the theory of convergence of probability distributions. The main point is that (22) implies (24). Similar reasoning shows that, conversely, (24) implies (22); you don't have to give the argument for that.

Exercise 10. Let $F_1, F_2, \ldots, F_n, \ldots$, and F be dfs with corresponding left-continuous inverse dfs $F_1^*, F_2^*, \ldots, F_n^*, \ldots$, and F^*. Suppose that

$$\lim_{n\to\infty} F_n(x) = F(x) \text{ for all continuity points } x \text{ of } F. \tag{22}$$

(a) Suppose that $u \in (0,1)$ and that w is a continuity point of F with $F^*(u) > w$. Use the switching formula to show that $F_n^*(u) > w$ for all large n.

(b) Suppose that $u \in (0,1)$ and that y is a continuity point of F with $F^*(u+) < y$. Show that $F_n^*(u) \le y$ for all large n.

(c) Use parts (a) and (b) of this exercise and part (b) of the preceeding exercise to show that for each $u \in (0,1)$, one has

$$F^*(u) \le \liminf_n F_n^*(u) \le \limsup_n F_n^*(u) \le F^*(u+). \tag{23}$$

(d) Use part (c) to show that

$$\lim_{n\to\infty} F_n^*(u) = F^*(u) \text{ for all continuity points } u \text{ of } F^*. \tag{24}$$

(e) Show by example that if u is not a continuity point of F^*, then $F_n^*(u)$ may not converge as $n \to \infty$. $\diamond$