Chapter 7

Distribution and Quantile Functions

1 Character of Distribution Functions

Let $X : (\Omega, \mathcal{A}, P) \to (R, \mathcal{B})$ denote a rv with *distribution function (df)* F_X, where

(1) $\qquad F_X(x) \equiv P(X \le x) \qquad \text{for } -\infty < x < \infty.$

Then $F \equiv F_X$ was seen earlier to satisfy

(2) $\qquad F \text{ is } \nearrow \text{ and right continuous, with } F(-\infty) = 0 \text{ and } F(+\infty) = 1.$

Because of the following proposition, any function F satisfying (2) will be called a df. [If F is $\nearrow$, right continuous, $0 \le F(-\infty)$, and $F(+\infty) \le 1$, we earlier agreed to call F a *sub-df.* As usual, $F(a, b] \equiv F(b) - F(a)$ denotes the increments of F, and $\Delta F(x) \equiv F(x) - F_-(x)$ is the mass of F at x.]

Proposition 1.1 (There exists an X with df F) If F satisfies (2), then there exists a probability space $(\Omega, \mathcal{A}, P)$ and a rv $X : (\Omega, \mathcal{A}, P) \to (R, \mathcal{B})$ for which the df of X is F. We write $X \cong F$.

Proof. The corollary to the correspondence theorem (theorem 1.3.1) shows that there exists a unique probability distribution P on $(\Omega, \mathcal{A}) \equiv (R, \mathcal{B})$ for which $P((a, b]) = F(b) - F(a)$ for all $a \le b$. Now define $X(\omega) = \omega$ for all $\omega \in R$ to be the identity function on R. $\square$

Theorem 1.1 (Decomposition of a df) Any df F can be decomposed as

(3) $\qquad F = F_d + F_c = F_d + F_s + F_{ac} = (F_d + F_s) + F_{ac}\,,$

where F_d, F_c, F_s, and F_{ac} are the unique sub-dfs of the following types (unique among those sub-dfs equal to 0 at $-\infty$):

(4) $\qquad F_d$ is a step function of the form $\sum_j b_j 1_{[a_j,\infty)}$ (with all $b_j > 0$).

(5) F_c is continuous.

(6) F_s is singular, with its measure orthogonal to Lebesgue measure.

(7) $F_{ac} = \int_{-\infty}^{\cdot} f_{ac}(y)\, dy$
for some $f_{ac} \geq 0$ that is finite, measurable, and unique a.e. λ.

Proof. Let $\{a_j\}$ denote the set of all discontinuities of F, which can only be jumps; and let $b_j \equiv F(a_j) - F_-(a_j)$. There can be only a countable number of jumps, since the number of jumps of size exceeding size $1/n$ is certainly bounded by n. Now define $F_d \equiv \sum_j b_j 1_{[a_j,\infty)}$, which is obviously $\nearrow$ and right continuous, since $F_d(x,y] \leq F(x,y] \searrow 0$ as $y \searrow x$ (the inequality holds, since the sum of jump sizes over every finite number of jumps between a and b is clearly bounded by $F(x,y]$, and then just pass to the limit). Define $F_c = F - F_d$. Now, F_c is $\nearrow$, since for $x \leq y$ we have $F_c(x,y] = F(x,y] - F_d(x,y] \geq 0$. Now, F_c is the difference of right-continuous functions, and hence is right continuous; it is left continuous, since for $x \nearrow y$ we have

(a) $F_c(x,y] = F(x,y] - \sum_{x<a_j\leq y} b_j = F_-(y) - F(x) - \sum_{x<a_j<y} b_j \to 0 - 0 = 0.$

We turn to the uniqueness of F_d. Assume that $F_c + F_d = F = G_c + G_d$ for some other $G_d \equiv \sum_j \bar{b}_j 1_{[\bar{a}_j,\infty)}$ with distinct $\bar{a}_j$'s and $\sum_j \bar{b}_j \leq 1$. Then $F_d - G_d = G_c - F_c$ is continuous. If $G_d \neq F_d$, then either some jump point or some jump size disagrees. No matter which disagrees, at some a we must have

(b) $\Delta F_d(a) - \Delta G_d(a) \neq 0,$

contradicting the continuity of $G_c - F_c = F_d - G_d$. Thus $G_d = F_d$, and hence $F_c = G_c$. This completes the first decomposition.

We now turn to the further decomposition of F_c. Associate a measure μ_c with F_c via $\mu_c((-\infty,x]) = F_c(x)$. Then the Lebesgue decomposition theorem shows that $\mu_c = \mu_s + \mu_{ac}$, where $\mu_s(B) = 0$ and $\mu_{ac}(B^c) = 0$ for some $B \in \mathcal{B}$; we say that μ_s and μ_{ac} are singular. Moreover, this same Lebesgue theorem implies the claimed uniqueness and shows that f_{ac} exists with the uniqueness claimed. Now, $F_{ac}(x) \equiv \mu_{ac}((-\infty,x]) = \int_{-\infty}^{x} f_{ac}(y)\, dy$ is continuous by $F_{ac}(x,y] \leq \mu_{ac}(x,y] \to 0$ as $y \to x$ or as $x \to y$. Thus $F_s \equiv F_c - F_{ac}$ is continuous, and $F_s(x) = \mu_s((-\infty,x])$. $\square$

Example 1.1 (The Lebesgue singular df) Define the *Cantor set* C by

(8) $C \equiv \{x \in [0,1] : x = \sum_{n=1}^{\infty} 2a_n/3^n, \text{ with all } a_n \text{ equal to 0 or 1}\}.$

[Thus the Cantor set is obtained by removing from $[0,1]$ the open interval $(\frac{1}{3}, \frac{2}{3})$ at stage one, then the open intervals $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ at stage two, ...] Finally, we define F on C by

(9) $F(\sum_{n=1}^{\infty} 2a_n/3^n) = \sum_{n=1}^{\infty} a_n/2^n.$

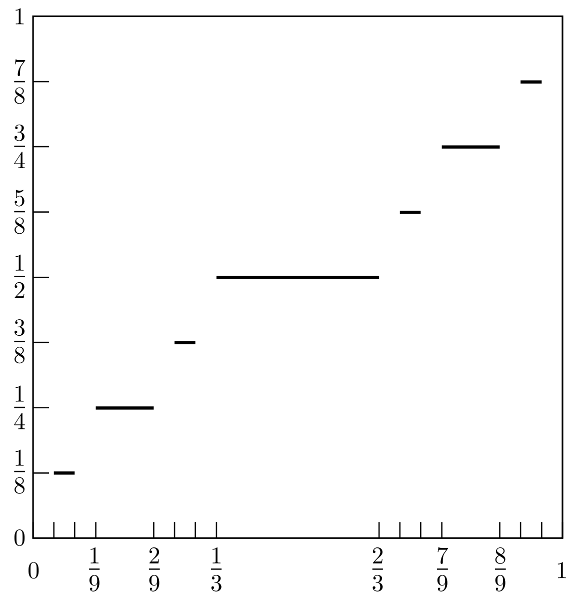

Figure 1.1 Lebesgue singular function.

Now note that $\{F(x) : x \in C\} = [0,1]$, since the right-hand side of (9) represents all of $[0,1]$ via dyadic expansion. We now define F linearly on C^c (the first three components are shown in figure 1.1 above). Since the resulting F is $\nearrow$ and achieves every value in $[0,1]$, it must be that F is continuous. Now, F assigns no mass to the "flat spots" whose lengths sums to 1 since $\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots = \frac{1/3}{1-2/3} = 1$. Thus F is singular with respect to Lebesgue measure λ, using $\lambda(C^c) = 1$ and $\mu_F(C^c) = 0$. We call this F the *Lebesgue singular df.* [The theorem in the next section shows that removing the flat spots does, even for a general df F, leave only the essentials.] □

Exercise 1.1 Let $X \cong N(0,1)$ (as in (9.1.22) below), and let $Y \equiv 2X$.
(a) Is the df $F(\cdot,\cdot)$ of (X,Y) continuous?
(b) Does the measure μ_F on R_2 have a density with respect to two-dimensional Lebesgue measure? [Hint. Appeal to corollary 2 to Fubini's theorem.]

Definition 1.1 Two rvs X and Y are said to be of the same *type* if $Y \cong aX + b$ for some $a > 0$. Their dfs are also said to be of the same type.

2 Properties of Distribution Functions

Definition 2.1 The *support* of a given df $F \equiv F_X$ is defined to be the minimal closed set C having $P(X \in C) = 1$. A point x is a *point of increase* of F if every open interval U containing x has $P(X \in U) > 0$. A *realizable t-quantile of F*, for $0 < t < 1$, is any value z for which $F(z) = t$.

Theorem 2.1 (Jumps and flat spots) Let $0 \le t \le 1$. Let U_t denote the maximal open interval of x's for which $F(x) = t$. The set of t's for which these U_t's are nonvoid is at most countable (these are exactly those t's that have more than one realizable t-quantile). Moreover:
(a) $C \equiv (\bigcup_{0\le t\le 1} U_t)^c$ is a closed set having $P(C) = 1$.
(b) C is equal to the set of all points of increase.
(c) C is the support of F.
(d) F has at most a countable number of discontinuities, and these discontinuities are all discontinuities of the jump type.
(e) F has an at most countable number of flat spots (the nonvoid U_t's).
[We will denote jump points and jump sizes of F by a_i's and b_i's. The t values and the $\lambda(U_t)$ values of the multiply realizable t-quantiles will be seen in the proof of proposition 7.3.1 below to correspond to the jump points c_j and the jump values d_j of the function $K(\cdot) \equiv F^{-1}(\cdot)$, and there at most countably many of them.]

Proof. (a) For each t there is a maximal open set U_t (possibly void) on which F equals t. Now, $P(X \in U_t) = 0$ using proposition 1.1.2. Note that $C \equiv (\cup_t U_t)^c$ is closed (since the union of an arbitrary collection of open sets is open). Hence $C^c = \cup_{0<t<1} U_t = \cup(a_n, b_n)$, where $(a_1, b_1), \ldots$ are (at most countably many) disjoint open intervals, and all those with $0 < t < 1$ must be finite. Now, by proposition 1.1.2, for the finite intervals we have $P(X \in (a_n, b_n)) = \lim_{\epsilon\to 0} P(X \in [a_n + \epsilon, b_n - \epsilon]) = \lim_{\epsilon\to 0} 0 = 0$, whence $P(X \in [a_n + \epsilon, b_n - \epsilon]) = 0$ follows, since this finite closed interval must have a finite subcover by U_t sets. If $(a_n, b_n) = (-\infty, b_n)$, then $P(X \in (-\infty, b_n)) = 0$, since $P(X \in [-1/\epsilon, b_n - \epsilon]) = 0$ as before. An analogous argument works if $(a_n, b_n) = (a_n, \infty)$. Thus $P(X \in C^c) = 0$ and $P(X \in C) = 1$. Note that the U_t's are just the (a_n, b_n)'s in disguise; each $U_t \subset$ some (a_n, b_n), and hence $U_t = \text{that}(a_n, b_n)$. Thus U_t is nonvoid for at most countably many t's.

(b) Let $x \in C$ and let U denote a neighborhood of x. Let $t \equiv F(x)$ and assume $P(U) = 0$. Then $x \in U \subset U_t \subset C^c$, which is a contradiction of $x \in C$. Thus all points $x \in C$ are points of increase. Now suppose conversely that x is a point of increase. Assume $x \notin C$. Then $x \in$ some (a_n, b_n) having $P(X \in (a_n, b_n)) = 0$, which is a contradiction. Thus $x \in C$. Thus the closed set C is exactly the set of points of increase, and $P(X \in C) = 1$.

(c) Assume that C is not the minimal closed set having probability 1. Then $P(\tilde{C}) = 1$ for some closed $\tilde{C} \subset C$. Let $x \in C \setminus \tilde{C}$ and let $t = F(x)$. Since $\tilde{C}^c$ is open, there exists an open set V_x with $x \in V_x \subset \tilde{C}^c$ and $P(X \in V_x) = 0$. Thus $x \in V_x \subset U_t \subset C^c$. Thus $x \notin C$, which is a contradiction. Thus C is minimal. □

3 The Quantile Transformation

Definition 3.1 (Quantile function) For any df $F(\cdot)$ we define the *quantile function (qf)* (which is the inverse of the df) by

(1) $$K(t) \equiv F^{-1}(t) \equiv \inf\{x : F(x) \geq t\} \qquad \text{for } 0 < t < 1.$$

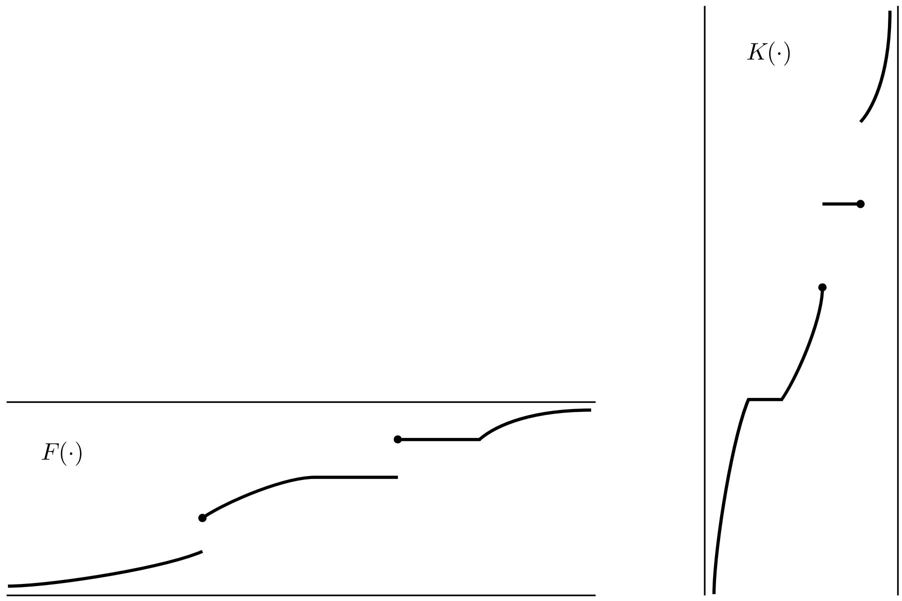

Figure 3.1 The df $F(\cdot)$ and the qf $K(\cdot) = F^{-1}(\cdot)$.

Theorem 3.1 (The inverse transformation) Let

(2) $$X \equiv K(\xi) \equiv F^{-1}(\xi), \qquad \text{where } \xi \cong \text{Uniform}(0,1).$$

Then

(3) $$[X \leq x] = [\xi \leq F(x)] \qquad \text{for every real } x,$$

(4) $$1_{[X \leq \cdot]} = 1_{[\xi \leq F(\cdot)]} \qquad \text{on } R, \text{ for every } \omega,$$

(5) $$X \equiv K(\xi) \equiv F^{-1}(\xi) \qquad \text{has df } F\ ,$$

(6) $$1_{[X < \cdot]} = 1_{[\xi < F_-(\cdot)]} \qquad \text{on } R, \text{ for a.e. } \omega.$$

Proof. Now, $\xi \le F(x)$ implies $X = F^{-1}(\xi) \le x$ by (1). If $X = F^{-1}(\xi) \le x$, then $F(x+\epsilon) \ge \xi$ for all $\epsilon > 0$; so right continuity implies $F(x) \ge \xi$. Thus (3) holds; (4) and (5) are then immediate.

If $\xi(\omega) = t$ where t is not in the range of F, then (6) holds. If $\xi(\omega) = t$ where $F(x) = t$ for exactly one x, then (6) holds. If $\xi(\omega) = t$ where $F(x) = t$ for at least two distinct x's, then (6) fails; theorem 7.2.1 shows that this can happen for at most a countable number of t's. (Or: Graph a df F that exhibits the three types of points t, and the rest is trivial with respect to (6), since the value of F at any other point is immaterial. Specifically, (6) holds for ω unless F has a flat spot at height $t \equiv F(\xi(\omega))$. Note figure 3.1.) □

Definition 3.2 (Convergence in quantile) Let K_n denote the qf associated with df F_n, for each $n \ge 0$. We write $K_n \to_d K_0$ to mean that $K_n(t) \to K_0(t)$ at each continuity point t of K_0 in $(0, 1)$. We then say that K_n *converges in quantile* to K_0, or K_n *converges in distribution* to K.

Proposition 3.1 (Convergence in distribution equals convergence in quantile)

$$(7) \qquad F_n \to_d F \qquad \text{if and only if} \qquad K_n \to_d K.$$

Proof. Suppose $F_n \to_d F$. Let $t \in (0, 1)$ be such that there is at most one value x having $F(x) = t$ (that is, there is not a multiply realizable t-quantile). Let $z \equiv F^{-1}(t)$.

First: We have $F(x) < t$ for $x < z$. Thus $F_n(x) < t$ for $n \ge$ (some N_x), provided that $x < z$ is a continuity point of F. Thus $F_n^{-1}(t) \ge x$ for $n \ge N_x$, provided that $x < z$ is a continuity point of F. Thus $\liminf F_n^{-1}(t) \ge x$, provided that $x < z$ is a continuity point of F. Thus $\liminf F_n^{-1}(t) \ge z$, since there are continuity points x that $\nearrow z$. Second: We also have $F(x) > t$ for $z < x$. Thus $F_n(x) > t$, and hence $F_n^{-1}(t) \le x$ for $n \ge$ (some N_x), provided that $x > z$ is a continuity point of F. Thus $\limsup F_n^{-1}(t) \le x$, provided that $x > z$ is a continuity point of F. Thus $\limsup F_n^{-1}(t) \le z$, since there are continuity points x that $\searrow z$.

Summary $F_n^{-1}(t) \to F^{-1}(t)$ for all but at most a countably infinite number of t's (namely, for all but those t's that have multiply realizable t-quantiles; these correspond to the heights of flat spots of F, and these flat spot heights t are exactly the discontinuity points of K). That is, $K_n \to_d K$.

The proof of the converse is virtually identical. □

Exercise 3.1 (Left continuity of K) Show that $K(t) = F^{-1}(t)$ is left continuous on $(0, 1)$. [Note that K is discontinuous at $t \in (0, 1)$ if and only if the corresponding U_t is nonvoid (see theorem 7.2.1). Likewise, the jump points c_j and the jump sizes d_j of $K(\cdot)$ are equal to the t values and the $\lambda(U_t)$ values of the multiply realizable t-quantiles.] [We earlier agreed to use a_i and b_i for the jump points and jump sizes of the associated df F.]

Exercise 3.2 (Properties of dfs) (i) For any df F we have

$$F \circ F^{-1}(t) \geq t \qquad \text{for all } 0 \leq t \leq 1,$$

and equality fails if and only if $t \in (0,1)$ is not in the range of F on $[-\infty, \infty]$.
(ii) (The probability integral transformation) If X has a continuous df F, then $F(X) \cong \text{Uniform}(0,1)$. In fact, for any df F,

$$P(F(X) \leq t) \leq t \qquad \text{for all } 0 \leq t \leq 1\,,$$

with equality failing if and only if t is not in the closure of the range of F.
(iii) For any df F we have

$$F^{-1} \circ F(x) \leq x \qquad \text{for all } -\infty < x < \infty\,,$$

and equality fails if and only if $F(y) = F(x)$ for some $y < x$. Thus

$$P(F^{-1} \circ F(X) \neq X) = 0 \qquad \text{whenever } X \cong F.$$

(iv) If F is a continuous df and $F(X) \cong \text{Uniform}(0,1)$, then $X \cong F$.

Proposition 3.2 (The randomized probability integral transformation) Let X denote an arbitrary rv. Let F denote its df, and let (a_j, b_j)'s denote an enumeration of whatever pairs (jump point, jump size) the df F possesses. Let $\eta_1, \eta_2, \ldots$ denote iid Uniform$(0,1)$ rvs (that are also independent of X). Then

(8) $$\dot{\xi} \equiv F(X) - \textstyle\sum_j b_j \eta_j \, 1_{[X=a_j]} \cong \text{Uniform}(0,1),$$

(9) $$X = F^{-1}(\dot{\xi}) = K(\xi).$$

[We have reproduced X from a Uniform$(0,1)$ rv that was defined using both X and some independent extraneous variation. Note figure 3.1.]

Proof. We have merely smoothed out the mass b_j that $F(X)$ placed at $F(a_j)$ by subtracting the random fractional amount $\eta_j b_j$ of the mass b_j. □

Exercise 3.3 (Change of variable) Suppose that rvs $X \cong F$ and $Y \cong G$ are related by $G(H) = F$ and $X = H^{-1}(Y)$, where H is right continuous on the real line with left-continuous inverse H^{-1}. (a) Then set g, X, μ, μ_X, A' in the theorem of the unconscious statistician equal to $g, H^{-1}, G, F, (-\infty, x]$ to conclude that

(10) $$\int_{(-\infty, H(x)]} g(H^{-1}) \, dG = \int_{(-\infty, x]} g \, dF,$$

since $(H^{-1})^{-1}((-\infty, x]) = \{y : H^{-1}(y) \leq x\} = (-\infty, H(x)]$.
(b) Making the identifications $G = I$, $H = F$, and $Y = \xi \cong \text{Uniform}(0,1)$ gives especially (via part (a), or via (2) and (3))

(11) $$\int_{[0, F(x)]} g(F^{-1}(t)) \, dt = \int_{(-\infty, x]} g \, dF$$

for arbitrary df F and any measurable g.

Proof. We now prove proposition 1.2.3. Let D be a subset of $[0,1]$ that is not Lebesgue measurable; its existence is guaranteed by proposition 1.2.2. Let $B \equiv F^{-1}(D)$. Then B is a subset of the Cantor set C of example 7.1.1. Since $\lambda(C) = 0$ and $B \subset C$, it follows that B is a Lebesgue set with $\lambda(B) = 0$. We now assume that B is Borel set (and look for a contradiction). We note that $F(B)$ is also a Borel set, since F (being $\nearrow$) is Borel measurable. However, F is 1-to-1 on C, and so $F(B) = D$. That is, D is a Borel set, and hence D is a Lebesgue set. This is the contradiction we sought. □

The Elementary Skorokhod Construction Theorem

Let $X_0, X_1, X_2, \ldots$ be iid F. Then $X_n \to_d X_0$, but the X_n do not converge to X_0 in the sense of $\to_{a.s.}$, $\to_p$, or $\to_r$. However, when general $X_n \to_d X_0$, it is possible to replace the X_n's by rvs Y_n having the same (marginal) dfs, for which the stronger result $Y_n \to_{a.s.} Y_0$ holds.

Theorem 3.2 (Skorokhod) Suppose that $X_n \to_d X_0$. Define $\xi(\omega) = \omega$ for each $\omega \in [0,1]$ so that $\xi \cong \text{Uniform}(0,1)$ on $(\Omega, \mathcal{A}, P) \equiv ([0,1], \mathcal{B} \cap [0,1], \lambda)$, for Lebesgue measure λ. Let F_n denote the df of X_n, and define $Y_n \equiv F_n^{-1}(\xi)$ for all $n \geq 0$. Let D_{K_0} denote the at most countable discontinuity set of K_0. Then both

$$
\begin{aligned}
&Y_n \equiv K_n(\xi) \equiv F_n^{-1}(\xi) \cong X_n \cong F_n \quad \text{for all } n \geq 0 \qquad \text{and} \\
&Y_n(\omega) \to Y_0(\omega) \quad \text{for all } \omega \notin D_{K_0}.
\end{aligned} \tag{12}
$$

Proof. This follows trivially from proposition 3.1. □

Exercise 3.4 (Wasserstein distance) For $k = 1$ or 2, define

$$\mathcal{F}_k \equiv \{F : F \text{ is a df, and } \textstyle\int |x|^k \, dF(x) < \infty\},$$

$$d_k(F_1, F_2) \equiv \int_0^1 |F_1^{-1}(t) - F_2^{-1}(t)|^k \, dt \qquad \text{for all } F_1, F_2 \in \mathcal{F}_k\,.$$

Show that both $(\mathcal{F}_k, d_k)$ spaces are complete metric spaces, and that

$$
\begin{aligned}
&d_k(F_n, F_0) \to 0 \qquad (\text{with all } \{F_n\}_0^\infty \in \mathcal{F}_k) \qquad \text{if and only if} \\
&F_n \to_d F_0 \quad \text{and} \quad \textstyle\int |x|^k \, dF_n(x) \to \int |x|^k \, dF_0(x)\,.
\end{aligned} \tag{13}
$$