Errata for
Probability, Random Processes, and Ergodic Properties: Second Edition

Robert M. Gray

Formatted April 3, 2023

1 Introduction

This document collects errata of the Second Edition of *Probability, Random Processes, and Ergodic Properties* with occasional reference to the free updated hard cover original First Edition published in 1988.

The most serious error in the book was discovered by Tamas Linder based on drafts of the Second Edition. The error originated in the First Edition but unfortunately was not caught until the Second Edition was already in press. These notes are intended to correct the faulty results and to also correct several minor typographical errors discovered by the author and readers since.

The proof of Lemma 9.2 on page 266 (Lemma 8.8.2 in the First Edition) is incorrect and the lemma is not stated accurately. In addition, the lemma does not hold in the implied generality and hence its use to prove Theorem 9.2 part (c) (Theorem 8.3.1 (d) in the first addition) and part (f) (Theorem 8.3.1 part (g) in the First Edition) is not justified. A correct proof of Theorem 9.2 (c) is provided.

The original version of these errata (dated 19 October 2010) provided the primary correction and a few others. This update incorporates several additional corrections.

This document was updated during March 2023 to include corrections from Jun Muramatsu of NTT. While correcting the typos he pointed out I found a few other typos in the book and email archives and the original errata list that are corrected here.

2 Lemma 9.2 & Implications

This section provides a corrected version of Lemma 9.2 and a corrected proof of Theorem 9.2 (c) which follows the original paper [R. M. Gray, D. L. Neuhoff and P. C. Shields, "A generalization of Ornstein's d–bar distance with applications to information theory," *Annals of Probability*, Vol. 3, No. 2, pp. 315–328, April 1975] .

The proof of Theorem 9.2 (f) given on p. 285 (mislabeled as (g)) is not correct because it is based on the incorrect original Lemma 9.2. The result is discussed but not proved here.

The corrected statement of the Lemma is given next. The key change is that "$\mathcal{G}$ is a standard generating field" replaces the weaker assumption "$\mathcal{G}$ is a countable generating field;" that is, the lemma holds specifically for the countable generating field formed as the union of the finite fields constituting a basis. The conclusions need not hold for an arbitrary countable generating field.

Lemma 9.2

Assume that $(\Omega, \mathcal{B})$ is standard and $\mathcal{G}$ is a standard generating field. Then $(\mathcal{P}(\Omega, \mathcal{B}), d_{\mathcal{G}})$ is sequentially compact; that is, if $\{\mu_n\}$ is a sequence of probability measures in $\mathcal{P}(\Omega, \mathcal{B})$, then it has a subsequence μ_{n_k} that converges.

Proof: Suppose that $\{\mu_n\}$ is a sequence of probability measures in $\mathcal{P}(\Omega, \mathcal{B})$. Since $\mathcal{G}$ is countable, we can use the standard (Cantor) diagonalization procedure to extract a subsequence μ_{n_k} such that $\lim_{k\to\infty} \mu_{n_k}(G)$ converges for all $G \in \mathcal{G}$. In particular, if $\mathcal{G} = \{G_i; i = 0, 1, \ldots\}$, first pick a subsequence of n for which $\mu_n(G_0)$ converges, then pick a further subsequence for which $\mu_n(G_1)$ converges, and so on. The result is a set function $\lambda(G)$ defined for $G \in \mathcal{G}$. This set function is obviously nonnegative and normalized $\lambda(\Omega) = 1$. Furthermore, λ is finitely additive on $\mathcal{G}$. This follows since $\mathcal{G} = \bigcup \mathcal{F}_n$, the union of a collection of finite fields (the basis) and hence given any two disjoint sets $F, G \in \mathcal{G}$, there must be some finite N for which $F, G \in \mathcal{F}_n$ for all $n \geq N$ and hence $\mu_n(F \cup G) = \mu_n(F) + \mu_n(G)$ for all $n \geq N$ so that $\lambda(F \cup G) = \lambda(F) + \lambda(G)$. Since the field $\mathcal{G}$ is standard, there is a unique extension of the finitely additive set function λ to a countably additive set function on $\mathcal{G}$, which in turn has an extension to a probability measure, say μ, on $\sigma(\mathcal{G})$ from the Caratheodory extension theorem. By construction, $\lim_{n\to\infty} \mu_n(G) = \mu(G)$ for all $G \in \mathcal{G}$ and hence $\lim_{n\to\infty} d_{\mathcal{G}}(\mu_n, \mu) = 0$. □

The example following Lemma 9.2 relates to Lemma 9.1 and not Lemma 9.2 since in the example $\mathcal{G}$ is a countable generating field as required by the first part of Lemma 9.1, but it is not standard as required by Lemma 9.2. To clarify this, the title of the subsection entitled "An Example" should be Distributional Distance and Weak Convergence" and the first part of the first sentence should be changed from "As an example of the previous construction" to "As an example of the implications of convergence with respect to distributional distance".

Theorem 9.2.(c) Proof

For completeness the entire proof is given rather than weaving in corrections. The corrected proof follows as closely as possible the structure and notation of the incorrect proof.

Given $\epsilon > 0$ let $\pi \in \mathcal{P}_s(\mu_X, \mu_Y)$ be such that $E_\pi \rho_1(X_0, Y_0) \leq \overline{\rho}'(\mu_X, \mu_Y) + \epsilon$. The induced distribution on $\{X^n, Y^n\}$ is then contained in $\mathcal{P}_n$, and hence using the stationarity of the processes

$$\overline{\rho}_n(\mu_{X^n}, \mu_{Y^n}) \leq E\rho_n(X^n, Y^n) = nE\rho_1(X_0, Y_0) \leq n(\overline{\rho}'(\mu_X, \mu_Y) + \epsilon),$$

and therefore normalizing by n and taking the limit implies $\overline{\rho}' \geq \overline{\rho}$ since ϵ is arbitrary.

Let $\pi^n \in \mathcal{P}_n,\ n = 1, 2, \dots$ be a sequence of measures such that

$$E_{\pi^n}[\rho_n(X^n, Y^n)] \leq \overline{\rho}_n + \epsilon_n$$

where $\epsilon_n \to 0$ as $n \to \infty$. Let q_n denote the product (block independent) measure on the sequence space $(A^{\mathbb{I}}, \mathcal{B}^{\mathbb{I}})^2$ induced by the π_n as explained next.

Let $\mathcal{G}$ denote a countable generating field for the standard space $(A^{\mathbb{I}}, \mathcal{B}^{\mathbb{I}})$. For any N and N-dimensional rectangle $F = \times_{i\in\mathbb{I}} F_i$ with all but a finite number N of the F_i being A^2 and the remainder being in $\mathcal{G}^2$

$$q_n(F) = \prod_{j\in\mathbb{I}} \pi^n(\times_{i=jn}^{(j+1)n-1} F_i).$$

Thus q_n is the pair process distribution obtained by gluing together independent copies of π^n. This measure is n-stationary by construction and we form a stationary process distribution π_n by averaging over n-shifts as

$$\pi_n(F) = \frac{1}{n}\sum_{i=0}^{n-1} q_n(T^{-i}F)$$

for all events F.

This measure on the rectangles extends to a stationary pair process distribution π_n on the sequence space $(A^{\mathbb{I}}, \mathcal{B}^{\mathbb{I}})^2$. For any $m = 1, 2, \dots, n$ the mth marginal restrictions of the π_n can be related to corresponding original marginal distributions. For example, consider the Y marginal and let $G = \times_{i=0}^{m-1} G_i \in \mathcal{G}^m$. Then

$$\begin{aligned} q_n(\ x, y : x^m \in A^m, y^m \in G\}) &= \pi_n(A^n \times (G \times A^{n-m})) \\ &= \mu_{Y^n}(G \times A^{n-m}) = \mu_{Y^m}(G). \end{aligned} \tag{1}$$

Similarly

$$q_n(\ x, y : x^m \in G, y^m \in A^m\}) = \mu_{X^m}(G).$$

Thus

$$\begin{aligned} \pi_n^m(A^m \times G) &= \pi_n(\{(x, y) : x^m \in A^m, y^m \in \mathcal{G}) \\ &= \frac{n-m+1}{n}\mu_{Y^m}(G) + \frac{1}{n}\sum_{i=1}^{m-1} \mu_{Y^{m-i}}(\times_{k=i}^{m-i})\mu_{Y^i}(\times_{k=0}^{i-1} G_k) \end{aligned}$$

with a similar expression for $G \times A^m$.

Since there are a countable number of finite dimensional rectangles in $\mathcal{B}^{\mathbb{I}}$ with coordinates in $\mathcal{G}$, we can use a diagonalization argument to extract a subsesquence π_{n_k} of π_n which converges on all of the rectangles. To do this enumerate all the rectangles, then pick a subsequence converging on the first, then a further subsequence converging on the second, and so on. The result is a limiting measure π on the finite-dimensional rectangles, and this can be extended to a measure also denoted by π on $(A^{\mathbb{I}}, \mathcal{B}^{\mathbb{I}})^2$; that is, to a stationary pair process distribution. Eq. (1) implies that for each fixed m

$$\lim_{n\to\infty} \pi_n(A^m \times G) = \pi^m(A^m \times G) = \mu_{Y^m}G)$$
$$\lim_{n\to\infty} \pi_n(G \times A^m) = \pi^m(G \times A^m) = \mu_{X^m}G)$$

and hence for any cylinder $F \in \mathcal{B}$ that

$$\pi(A^{\mathbb{I}} \times F) = \mu_Y(F)$$
$$\pi(F \times A^{\mathbb{I}}) = \mu_X(F).$$

Thus π induces the desired marginals and hence $\pi \in \mathcal{P}_s$ and that

$$\begin{aligned}
\overline{\rho}'(\mu_X, \mu_Y) &\leq E_\pi \rho_1(X_0, Y_0) \\
&= \lim_{k\to\infty} E_{\pi_{n_k}} \rho_1(X_0, Y_0) \\
&= \lim_{k\to\infty} n_k^{-1} \sum_{i=0}^{n_k-1} E_{q_{n_k}} \rho_1(X_i, Y_i) \\
&= \lim_{k\to\infty} (\overline{\rho}_{n_k} + \epsilon_{n_k}) = \overline{\rho}(\mu_X, \mu_Y)
\end{aligned}$$

proving that $\overline{\rho}' \leq \overline{\rho}$ and hence that they are equal.

Theorem 9.2 (f)

The proof of Theorem 9.2 (f) (Theorem 8.3.1 in the First Edition) is incorrect because it used Lemma 9.2 (Lemma 8.2.2) in a situation where that lemma does not apply. In particular, the repaired Lemma 9.2 requires a standard generating field $\mathcal{G}$ and not just a countable generating field. Convergence in distributional distance with respect to a countable generating field is not sufficient to ensure that the average distortion between the coordinates of the limiting process is bound above by the limit supremum of the average distortions of the p_n. This is an upper semicontinuity result and the distributional distance with respect to the standard field required by Lemma 9.2 is not strong enough to guarantee it. This error is not contained in the cited references where the other parts of the theorem are proved, it first appeared in the First Edition. The fact that the infimum is a minimum for $\bar{\rho}_n$ is well known in the optimal transport literature, e.g., Villani [110-11], but the proof involves considerably more machinery than developed in the book. Specifically, it invokes Prohorov's theorem to demonstrate that under the given assumptions, given a sequence p_n there exists a limit p to which p_n converges weakly. The weak convergence, the continuity of distortion with respect to the underlying metric, and the bounded average distortions lead to a proof. Convergence in the sense of a distributional distance using a standard generating field is not strong enough to ensure weak convergence. Convergence in the sense of a distributional distance with respect to the countable generating field using the open sets as in Lemmas 1.10 and 9.3 is sufficient to ensure weak convergence, but not sufficient to ensure a convergent subsequence. I have not found a simpler proof taking advantage of Lemma 9.2. This result has been removed from the corrected First Edition.

3 Ergodic Properties

In Chapter 7 many variations on a single typo of occurred where a relative frequency $< f >$ of a measurement f appeared only as f. All of these errors are collected in this section.

Section 7.1

Page 196, first displayed equation on page:

$$< f >_n = n^{-1} \sum_{i=0}^{n-1} f(T^i x).$$

should be

$$< f >_n (x) = n^{-1} \sum_{i=0}^{n-1} f(T^i x).$$

Correction thanks to thanks to Jun Muramatsu, NTT.

Page 197 first displayed equation: replace

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = f(x)$$

by

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = < f > (x)$$

Page 197 second displayed equation replace

$$| < f >_n - f | \underset{n\to\infty}{\to} 0.$$

by

$$| < f >_n - < f >| \underset{n\to\infty}{\to} 0.$$

Page 197 third displayed equation: replace

$$\begin{aligned} &| < fT >_n - < f > | \\ &= |(\frac{n+1}{n}) < f >_{n+1} -n^{-1}f - < f >| \\ &\leq |(\frac{n+1}{n}) < f >_{n+1} - < f >_{n+1}| + | < f >_{n+1} -f | +n^{-1} | f | \\ &\leq \frac{1}{n} | < f >_{n+1}| + | < f >_{n+1} -f | + \frac{1}{n}|f| \underset{n\to\infty}{\to} 0 \end{aligned}$$

where the middle term goes to zero by assumption, implying that the first term must also go to zero, and the right-most term goes to zero since f cannot assume ∞ as a value. by

$$\begin{aligned} &| < fT >_n - < f > | \\ &= |(\frac{n+1}{n}) < f >_{n+1} -n^{-1}f - < f >| \\ &= | < f >_{n+1} - < f > +n^{-1}(< f >_{n+1} -f)| \\ &\leq | < f >_{n+1} - < f > | + n^{-1} | < f >_{n+1} | + n^{-1}|f| \underset{n\to\infty}{\to} 0 \end{aligned}$$

where the leftmost term goes to zero by assumption, the middle term goes to zero since $< f >_{n+1}$ converges to $< f >$, and the right-most term goes to zero since f cannot assume ∞ as a value.

Section 7.2

Page 200, Displayed equation above Lemma 7.3. Replace

$$\begin{aligned} | n^{-1} \sum_{i=0}^{n-1} \int_G fT^i \, dm - \int_G f \, dm | &= | \int_G (n^{-1} \sum_{i=0}^{n-1} fT^i - f) dm| \\ &\leq \int_G |n^{-1} \sum_{i=0}^{n-1} fT^i - f | \, dm \\ &\leq \| < f >_n -f\|_1 \underset{n\to\infty}{\to} 0. \end{aligned}$$

by

$$\begin{aligned} | n^{-1} \sum_{i=0}^{n-1} \int_G fT^i \, dm - \int_G < f > \, dm | &= | \int_G (n^{-1} \sum_{i=0}^{n-1} fT^i - < f >) dm| \\ &\leq \int_G |n^{-1} \sum_{i=0}^{n-1} fT^i - < f >| \, dm \\ &\leq \| < f >_n - < f > \|_1 \underset{n\to\infty}{\to} 0. \end{aligned}$$

Page 200-201, Eqs. (7.1)-(7.2). Replace

$$\lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} \int_G fT^i \, dm = \int_G f \, dm, \text{ all } G \in \mathcal{B}.$$

Thus, for example, if we take G as the whole space

$$\lim_{n\to\infty} n^{-1}\sum_{i=0}^{n-1} E_m(fT^i) = E_m f.$$

by

$$\lim_{n\to\infty} n^{-1}\sum_{i=0}^{n-1}\int_G fT^i\,dm = \int_G < f > \,dm, \text{ all } G\in\mathcal{B}.$$

Thus, for example, if we take G as the whole space

$$\lim_{n\to\infty} n^{-1}\sum_{i=0}^{n-1} E_m(fT^i) = E_m < f > .$$

Page 201, Eqs. (7.3)-(7.4): replace

$$\begin{aligned}\lim_{n\to\infty} n^{-1}\sum_{i=0}^{n-1} m(T^{-i}F\cap G) &= \lim_{n\to\infty} n^{-1}\sum_{i=0}^{n-1}\int_G 1_F T^i dm \\ &= E_m(1_F 1_G), \text{ all } F,G\in\mathcal{B}.(7.3)\end{aligned}$$

For example, if G is the entire space than (7.3) becomes

$$\lim_{n\to\infty} n^{-1}\sum_{i=0}^{n-1} m(T^{-i}F) = E_m(1_F), \text{ all events } F.(7.4)$$

by

$$\begin{aligned}\lim_{n\to\infty} n^{-1}\sum_{i=0}^{n-1} m(T^{-i}F\cap G) &= \lim_{n\to\infty} n^{-1}\sum_{i=0}^{n-1}\int_G 1_F T^i dm \\ &= E_m(< 1_F > 1_G), \text{ all } F,G\in\mathcal{B}.(7.3)\end{aligned}$$

For example, if G is the entire space than (7.3) becomes

$$\lim_{n\to\infty} n^{-1}\sum_{i=0}^{n-1} m(T^{-i}F) = E_m(< 1_F >), \text{ all events } F.(7.4)$$

Section 7.5

Page 222: Proof of Lemma 7.2, first displayed equation: Replace

$$\lim_{n\to\infty} E_m(< f >_n) = E_m f.$$

by

$$\lim_{n\to\infty} E_m(< f >_n) = E_m < f > .$$

Page 222: Proof of Lemma 7.2, second displayed equation: Replace

$$E_m f = E_{\overline{m}} f.$$

by

$$E_m < f >= E_{\overline{m}} f.$$

Section 7.6

Page 225, Lemmas 7.11 and Corollary 7.10: Replace 2nd displayed equation on page

$$f = E_{\overline{m}}(f|\mathcal{I}).$$

by

$$< f >= E_{\overline{m}}(f|\mathcal{I}).$$

Similarly make same replacement in last displayed equation on page.

Section 7.7

p. 227, first displayed formula in statement of Lemma 7.14. Replace

$$\lim_{n\to\infty} < f >_n = E_m f = \lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} E_m f T^i.$$

by

$$\lim_{n\to\infty} < f >_n = E_m < f > = \lim_{n\to\infty} n^{-1} \sum_{i=0}^{n-1} E_m f T^i.$$

Miscellaneous Typographical Errors

Page 2, Eq. (1.1) should read $d(a, b) = 0$ if and only if $a = b$.

Thanks to Michael D. White for catching this typo.

On p. 207, displayed equation in the middle of he page: replace

$$f(x) = 1_F(x) \lim_{n\to\infty} \frac{1}{n} \sum_{n=0}^{n-1} f(T^i x).$$

by

$$f(x) = 1_F(x) \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x).$$

Chapter 9 Corrections thanks to Jun Muramatsu, NTT.

Page 264, first line of Lemma 9.1: Replace

The metric space $(\mathcal{P}((\Omega, \mathcal{B})), d_{\mathcal{G}})$ The same error occurs in the statement of Lemma 9.2 on p. 266, which error is fixed in the corrected statement of the Lemma previously given.

by

The metric space $(\mathcal{P}(\Omega, \mathcal{B}), d_{\mathcal{G}})$

Page 276, line 9: replace

$$| \mu_X(G) - \mu_Y(G) | \leq \mu_Y(B^c) - \mu_X(B^c$$

by

$$| \mu_X(G) - \mu_Y(G) | \leq \mu_Y(B^c) - \mu_X(B^c)$$

line 14: replace

$$| \mu_X(G) - \mu_Y(G) | = \sum_x |\mu_X(x) - \mu_Y(x)|$$

by

$$| \mu_X - \mu_Y | = \sum_x |\mu_X(x) - \mu_Y(x)|$$

Page 277, Eq. (9.12): replace

$$\frac{(\mu_X(x) - \mu_Y(x))1_B(x)(\mu_Y(x) - \mu_X(x))1_{B^c}(x).}{\mu_X(B) - \mu_Y(B)}$$

by

$$\frac{(\mu_X(x) - \mu_Y(x))1_B(x)(\mu_Y(y) - \mu_X(y))1_{B^c}(y)}{\mu_X(B) - \mu_Y(B)}.$$