Robert M. Gray

Errata for Probability, Random Processes, and Ergodic Properties

Second Edition

October 19, 2010

Springer

This document collects the errata of the Second Edition of *Probability, Random Processes, and Ergodic Properties.* As time permits those errors and typos inherited from the first edition published in 1988 and recently discovered are being incorporated into the free online version of the corrected First Edition available at `http://ee.stanford.edu/ gray/arp.html`.

The proof of Lemma 9.2 (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 part (d) in the First Edition) is not justified. These errata provide a corrected version of the Lemma and a 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] and does not use the incorrect Lemma 9.2.

The proof of Theorem 9.2 (g) is missing. It requires Prohorov's theorem and significant material not covered in the book. The statement should be omitted. The corresponding result for the transportation distance can be found, e.g., in Villani.

Lemma 9.2

Below follows a correct statement of the Lemma and a correct proof. The statement of the Lemma should read "$\mathcal{G}$ is a standard generating field" rather than "$\mathcal{G}$ 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 as in the previous lemma 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 $mathcalG$. 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 9.1and 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 Distibutional 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)

For completeness the entire proof of (c) is included.

(c) 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 $\mathcal{P}_n(\mu_{X^n}, \mu_{Y^n})$, 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 $\overline{\rho}' \geq \overline{\rho}$ since ϵ is arbitrary.

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

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

where $\epsilon_n > 0$ and $\epsilon_n \to 0$ as $n \to \infty$. Let q_n denote the product (block independent) measure $(A^{\mathbb{T}}, \mathcal{B}(A)^{\mathbb{T}})^2$ induced by the π^n as explained next. Let $\mathcal{G}$ denote a countable generating field for the standard space $(A, \mathcal{B}(A))$. For any N and N-dimensional rectangle or cylinder of the form $F = \times_{i\in\mathbb{T}} F_i$ with all but a finite number N of the F_i being A^2 and the remainder being in $\mathcal{G}^2$ define

$$q_n(F) = \prod_{j\in\mathbb{T}} \pi^n(F_{jn} \times F_{jn+1} \times \cdots \times F_{jn+n-1}).$$

Thus q_n assigns a probability to rectangles in a way that treats successive n-tuples as independent. Next "stationarize" q_n to form a measure on rectangles by averaging over n-shifts to form

$$\pi_n(F) = \frac{1}{n}\sum_{i=0}^{n-1} q_n(T^{-i}F) = \frac{1}{n}\sum_{i=0}^{n-1}\prod_{j\in\mathbb{T}} \pi^n(F_{jn+i}\times F_{jn+i+1}\times\cdots\times F_{jn+i+n-1}).$$

This measure on the rectangles extends to a stationary pair process distribution. For any $m = 1, 2, \ldots, n$ we can relate the mth marginal restrictions of π_n to the corresponding original marginals. For example, consider the Y marginal and let $G = \times_{k=0}^{m-1} G_i \in \mathcal{G}^m$. Then

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

and similarly if $G \in \mathcal{B}(A)^m$ then

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

Thus

$$\pi_n^m(A^m \times G) \tag{0.1}$$
$$= \pi_n(\{(x, y) : x^m \in A^m, y^m \in G\})$$
$$= \frac{n-m+1}{n}\mu_{Y^m}(G) + \frac{1}{n}\sum_{i=1}^{m-1} \mu_{Y^{m-i}}(\times_{k=i}^{m-i})\mu_{Y^i}(\times_{k=0}^{i-1} G_k) \tag{0.2}$$

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

Since there are a countable number of finite dimensional rectangles in $\mathcal{B}^{\mathbb{T}}$ with coordinates in $\mathcal{G}$, we can use a diagonalization argument to extract a subsequence 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 $(A, \mathcal{B}(A))^2$, that is, to a stationary pair process distribution. Eq. (0.2) implies that for each fixed m

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

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

$$\begin{aligned} \pi(A^{\mathbb{T}} \times F) &= \mu_Y(F) \\ \pi(F \times A^{\mathbb{T}}) &= \mu_X(F) \end{aligned}$$

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

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

Other errata:

- In Chapter 7 there are several occurrences of f which should be $< f >$ or (equivalently) $\hat{f}$, the limiting time average. I list examples below, but there may be others.
  p. 197,

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

should be

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

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

should be

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

p. 200

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

should read

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

Eq. (7.1)

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

should read

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

p. 201, (7.2)

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

should be

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

(7.3)

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

should be

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

(7.4)

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

should be

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

p. 222 The first equation in the proof of Corollary 7.2

$$\lim_{n\to\infty} E_m < f >= E_m f$$

should be

$$\lim_{n\to\infty} E_m < f >= E_m < f >$$

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

should be

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

p. 225 twice

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

should be

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

p. 227

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

should be

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

- p. 197, third equation from the top, 00 should be just 0.