Math 254B Lecture 2 Notes

Daniel Raban

April 3, 2019

1 Marginals, Laws of Large Numbers, and The Asymptotic Equipartition Property

1.1 Marginal distributions and coupling

Denote $M(X, \mathcal{B})$ as the space of finite signed measures on $(X, \mathcal{B})$ and $P(X, \mathcal{B})$ as the subcollection of probability measures. We may also drop $\mathcal{B}$ in the notation, assuming it to be the Borel σ-algebra.

Definition 1.1. If $\lambda = \varphi_* \mathbb{P} \in P(\prod_i X_i, \bigotimes_i \mathcal{B}_i)$ is a joint distribution, then the dsitrubtions $\mu_i = \pi_{i*}\lambda$ are called the **marginals** of μ_i. λ is called a **coupling** of $(\mu_i)_i$.

Proposition 1.1. *$(\varphi_i)_i$ are independent if and only if $\varphi_* \mathbb{P} = \bigotimes_{i \in I} \varphi_{i*}\mathbb{P}$.*

Remark 1.1. Product measures can be defined as usual, provided $\mu_i(X_i) = 1$ for all i.

Proposition 1.2. *If $f_1, \dots, f_m$ are independent, $\mathbb{R}$-valued random variablea with $f_i \in L^1(\mathbb{P})$ for all i, then $\mathbb{E}[f_1, \dots, f_m] = \prod_{i=1}^m \mathbb{E}[f_i]$.*

Proof. Let μ_i be the distribution of f_i. By independence, the joint distribution of the f_i is $\mu_1 \otimes \cdots \otimes \mu_m$. Using this joint distribution, the left hand side is

$$\int_{\mathbb{R}^m} x_1 \cdots x_m \, d(\mu_1 \otimes \cdots \otimes \mu_m),$$

and the right hand side is

$$\int_{\mathbb{R}} x_1 \, d\mu_1 \cdots \int_{\mathbb{R}} x_m \, d\mu_m.$$

Tonelli's theorem tells you that the left hand side with absolute values is integrable, so then Fubini's theorem tells you that the right hand side equals the left. $\square$

1.2 Laws of large numbers

Theorem 1.1. *If $f_1, f_2, \dots$ are $\mathbb{R}$-valued, iid random variables with all $f_i \in L^1(\mathbb{P})$, then*

$$\frac{1}{n}\sum_{i=1}^{n} f_i \to E[f_i].$$

in $\|\cdot\|_1$ and a.e.

We will not prove this yet. But instead, for now, let's prove the special case of convergence in $\|\cdot\|_2$ when each $f_i \in L^2(\mathbb{P})$.

Proof. First, write

$$\frac{1}{n}\sum_{i=1}^{n} f_i - \mathbb{E}[f_1] = \frac{1}{n}\sum_{i=1}^{n}(f_i - \mathbb{E}[f_i]).$$

So we can assume that $\mathbb{E}[f_i] = 0$. Now

$$\left\| \frac{1}{n}\sum_{i=1}^{n} f_i \right\|_2^2 = \frac{1}{n^2}\sum_{i=1}^{n} \|f_i\|_2^2 + \frac{1}{n^2}\sum_{i \neq j} \mathbb{E}[f_i f_j]$$

But if $i \neq j$, then $\mathbb{E}[f_i f_j] = \mathbb{E}[f_i]\,\mathbb{E}[f_j] = 0$.

$$\begin{aligned} &= \frac{1}{n^2}\sum_{i=1}^{n} \|f_i\|_2^2 \\ &= O(1/n). \end{aligned}$$

□

Remark 1.2. You can prove the full theorem from this special case by using the fact that L^2 is dense in L^1 (since simple functions are dense in all L^p).

Let $|\mathcal{X}| < \infty$ be a finite alphabet. Let $n \in \mathbb{N}$, $x \in \mathcal{X}^n$, and $a \in \mathcal{X}$.

Definition 1.2. Let $N(a \mid x) = |\{i = 1, \dots, n : x_i = a\}|$. For a fixed x,

$$\mathbb{P}_x(a) := \frac{N(a \mid x)}{n} \in P(\mathcal{X})$$

is called the **empirical distribution** of x.

Corollary 1.1. *Let $\alpha_1, \alpha_2, \dots$ be iid random variables taking values in $\mathcal{X}$ with common distribution $\mathbb{P}$. Then $\mathbb{P}_{(\alpha_1, \dots, \alpha_n)} \to \mathbb{P}$ in probability and a.s.*

Proof. For all $a \in \mathcal{X}$,

$$\mathbb{P}_{(\alpha_1,\ldots,\alpha_n)}(a) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{a\}}(\alpha_i) \xrightarrow{\text{LLN}} \mathbb{E}[\mathbb{1}_{\{a\}}(\alpha_1)] = \mathbb{P}(a). \qquad \square$$

Definition 1.3. Let $p \in P(\mathcal{X})$. The **ε-typical set for** p is

$$T_\varepsilon^{(n)}(p) = \{x \in \mathcal{X}^n : \|\mathbb{P}_x - p\| < \varepsilon\}.$$

Corollary 1.2. *For all $\varepsilon > 0$, $p^{\otimes n}(T_\varepsilon^{(n)}(p)) \xrightarrow{n\to\infty} 1$.*

1.3 Asymptotic equipartition and Shannon entropy

Let $\alpha_1, \alpha_2, \ldots$ be iid $\mathcal{X}$-valued random variables with common distribution p. We call $\mathcal{X}$ the **source**.

Theorem 1.2 (asymptotic equipartition property, Shannon)**.** *As $n \to \infty$ the random sequence $(\alpha_1, \alpha_2, \ldots) \in \mathcal{X}^{\mathbb{N}}$ a.s. satisfies*

$$p^{\otimes n}((\alpha_1, \ldots, \alpha_n)) = e^{-H(p)+o(n)}$$

as $n \to \infty$, where

$$H(p) = -\sum_{x\in\mathcal{X}} p(x)\log(p(x)).$$

Proof. By the law of large numbers,

$$\begin{aligned}
\frac{1}{n}\log(p^{\otimes n}((\alpha_1,\ldots,\alpha_n))) &= \frac{1}{n}\sum_{i=1}^{n}\log(p(\alpha_i)) \\
&\xrightarrow{\text{LLN}} \mathbb{E}[\log(p(\alpha_1))] \\
&= \sum_{x\in\mathcal{X}} p(x)\log(p(x)) \\
&= -H(p). \qquad \square
\end{aligned}$$

Definition 1.4. The entropy ε-typical set from p is

$$A_\varepsilon^{(n)}(p) = \{x \in \mathcal{X}^n : e^{-(H(p)+\varepsilon)n} < p^{\otimes n}(x) < e^{-(H(p)-\varepsilon)n}.$$

Corollary 1.3. *For all $\varepsilon > 0$,*

$$p^{\otimes n}(A_\varepsilon^{(n)}(p)) \xrightarrow{n\to\infty} 1.$$

Definition 1.5. The quantity $H(p)$ is called the **Shannon entropy** of p.

If α is a random variable with distribution p, we write $H(\alpha) := H(p)$.

Definition 1.6. Let $p \in P(\mathcal{X})$, and let $a \in (0, 1)$. The **a-covering number of p** is

$$\mathrm{cov}_a(p) := \min\{|F| : F \subseteq \mathcal{X}, p(F) > a\}.$$

Here is a corollary of the AEP:

Corollary 1.4. *For all $a \in (0, 1)$,*

$$\mathrm{cov}_a(p^{\otimes n}) = e^{H(p)n+o(n)}.$$

Proof. Homework. □

The idea of the proof is to first pretend that $p^{\otimes n}$ is uniform and then to see that the errors are not exponentially big.