PROBABILITY 2020
EXAM

INSTRUCTOR: GIOVANNI PISTONE

1. Bernoulli probability function

Let X be a finite set, $\#X = n$. We define a probability function on the finite set of parts $S = \mathcal{P}(X)$, $\#S = 2^n$. If $s \in S$, that is, $s \subset X$, denote by the same letter both the set and its indicator function, $s(x) = 1$ if $x \in s$, 0 otherwise. One can check that for each $p \in]0,1[$,

$$p(s) = \prod_{x \in X} p^{s(x)}(1-p)^{1-s(x)} = p^{\#s}(1-p)^{n-\#s} = (1-p)^n \left(\frac{p}{1-p}\right)^{\#s}$$

is a probability function. It is the *Bernoulli* probability function.

For the corresponding probability measure it holds $\mathbb{P}_p\left(\#s = k\right) = \binom{n}{k} p^k (1-p)^{n-k}$. The equation $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$ defines the *binomial* probability function on $\{0, 1, \dots, n\}$.

Let $X = X_1 \cup X_2$ be a partition of X. We have

$$\mathbb{P}_p\left(s \colon s \subset X_1\right) = \sum_{s \subset X_1} p^{\#s}(1-p)^{n-\#s} = \sum_{s \subset X_1} p^{\#s}(1-p)^{\#X_1 - \#s}(1-p)^{\#X - \#X_1} = (1-p)^{\#X_2} \ .$$

For each $s \subset X$ define $s_1 = s \cap X_1$ and $s_2 = s \cap X_2$. Write $n_1 = \#X_1$, $n_2 = \#X_2$. The Bernoully probability function is decomposed as

$$p(s) = (1-p)^{n_1} \left(\frac{p}{1-p}\right)^{\#s_1} \times (1-p)^{n_2} \left(\frac{p}{1-p}\right)^{\#s_2} = p_1(s_1) p_2(s_2) \ .$$

[See below]

2. Random Graph

Let V be a finite set of *vertices*, $\#V = v$. The set of *edges* is the set of all 2-subsets of V. The usual notation is $\mathcal{E}$, but here we use $X = \binom{V}{2}$, $\#X = \binom{v}{2}$, for coherence with the previous section. A *graph* g is a set of edges. Let us consider the sample space of all graphs $G = \mathcal{P}(X)$, $\#G = 2^{\binom{v}{2}}$ and the Bernoully probability,

$$p(g) = (1-p)^{\binom{v}{2}} \left(\frac{p}{1-p}\right)^{\#g} \ .$$

Date: March 31, 2020.

All graphs with the same number of edges have the same probability. Under this model, the probability that a graph has e edges, $0 \le e \le \binom{v}{2}$ is

$$\binom{\binom{v}{2}}{e}(1-p)^{\binom{v}{2}}\left(\frac{p}{1-p}\right)^{e} .$$

Let us decompose the vertex set into $V = V_1 \cup V_2$, with numbers $v = v_1 + v_2$. This produces a partition of the set of edges into 3 parts, $X = X_1 \cup X_2 \cup X_{12}$. X_1 are the edges in V_1, X_2 are the edges in V_2, X_{12} are the edges connecting V_1 and V_2. Note that

$$\#X = \binom{v_1+v_2}{2}, \#X_1 = \binom{v_1}{2}, \#X = \binom{v_2}{2}, \#X_{12} = v_1 v_2 .$$

Every graph is decomposed as $g = g_1 \cup g_2 \cup g_{12}$ and

$$p(g) = p_1(g_1) \times p_2(g_2) \times (1-p)^{v_1 v_2}\left(\frac{p}{1-p}\right)^{\#g_{12}} .$$

[To be continued]

3. Monotone classes

Let S be the sample space. A family $\mathcal{M}$ of subsets of S is a *monotone class* if for all infinite sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{M}$, if the sequence is non-increasing $A_n \supset A_{n+1}$ then $\lim_{n\to\infty} A_n = \cap_n A_n \in \mathcal{M}$, and, if the sequence is non-decreasing $A_n \subset A_{n+1}$ then $\lim_{n\to\infty} A_n = \cup_n A_n \in \mathcal{M}$. A σ-algebra is a monotone class.

If $\mathcal{B}$ is a field on S, then the σ-algebra generated by $\mathcal{B}$, $\sigma(\mathcal{B})$, coincides with the monotone class $m(\mathcal{B})$ generated by $\mathcal{B}$. See [1, § 1.4].

Consider the relation $\Phi\colon (A,B)$ on the parts of S which is true if, and only if,

$$A \cup B, A \setminus B, B \setminus A \in m(\mathcal{B}) .$$

For all A, $\Phi(A) = \{B \mid A \cup B, A \setminus B, B \setminus A \in m(\mathcal{B})\}$ is a monotone class. Moreover, if $A \in \mathcal{B}$, then $\Phi(A)$ contains $\mathcal{B}$, hence $m(\mathcal{B}) \subset \Phi(A)$ for all $A \in \mathcal{B}$. In turn, let $A \in m(\mathcal{B})$. As $\Phi(A) \supset \mathcal{B}$ and is a monotone class, then $\Phi(A)$ contains $m(\mathcal{B})$, hence $m(\mathcal{B})$ is a field and a monotone class. Finally, a field which is a monotone class is a σ-algebra.

[To be continued]

4. Connected random graphs

See [2, Example 9.18]. A graph is connected if all for each couple of vertices there is a connecting path. Each vertex belongs to a connected component. Let us count the number $C(n)$ of connected graphs with n vertices. Let v be a distinguished vertex of the graph g and let $C(g)$ be the connected component of g containing u. Now, $\#C(g) = k$ if, and only if, 1) there is a subset C of k vertices containing u where g is connected, and 2) there are not edges connecting C whith its complement. The number of graphs with n vertices such that 1) and 2) is

$$\binom{n-1}{k-1} C(k) 2^{\binom{n-k}{2}} , \quad k = 1, \ldots, n .$$

Please, here and below, check a simple example.

As

$$\sum_{k=1}^{n} \binom{n-1}{k-1} C(k) 2^{\binom{n-k}{2}} = 2^{\binom{n}{2}} \ ,$$

we have the recursion

$$C(n) = 2^{\binom{n}{2}} - \sum_{k=1}^{n-1} \binom{n-1}{k-1} C(k) 2^{\binom{n-k}{2}} \ .$$

Let p be the Bernoulli parameter in the random graph model and let $P(n)$ be the probability that a graph with n vertices is connected. Notice that each $P(n)$ is a probability on a different sample space. Given a graph g with vertices V, there exist a unique subset of vertices C such that the subgraph g_1 is connected and the remaining subgraph g_2 has no edges connecting C. That is, $g = g_1 + g_2$, where g_1 is connected and g_2 is generic. It follows that

$$\begin{aligned}
1 &= \sum_g (1-p)^{\binom{n}{2}} \left(\frac{p}{1-p}\right)^{\#g_1} \left(\frac{p}{1-p}\right)^{\#g_2} \\
&= \sum_{k=1}^{n} \sum_{v \in C \in \binom{V}{k}} \sum_{g_1} (1-p)^{\binom{n}{2}} \left(\frac{p}{1-p}\right)^{\#g_1} \sum_{g_2} \left(\frac{p}{1-p}\right)^{\#g_2} \\
&= \sum_{k=1}^{n} (1-p)^{\binom{n}{2} - \binom{n-k}{2}} \sum_{v \in C \in \binom{V}{k}} \sum_{g_1} \left(\frac{p}{1-p}\right)^{\#g_1} \\
&= \sum_{k=1}^{n} (1-p)^{\binom{n}{2} - \binom{n-k}{2}} \binom{n-1}{k-1} \sum_{g_1} \left(\frac{p}{1-p}\right)^{\#g_1} \\
&= \sum_{k=1}^{n} (1-p)^{\binom{n}{2} - \binom{n-k}{2} - \binom{k}{2}} \binom{n-1}{k-1} P(k) \\
&= \sum_{k=1}^{n} (1-p)^{k(n-k)} \binom{n-1}{k-1} P(k) \ .
\end{aligned}$$

The last term is $P(n)$, then

$$P(n) = 1 - \sum_{k=1}^{n-1} (1-p)^{k(n-k)} \binom{n-1}{k-1} P(k) \ .$$

5. Classes which are closed under intersection

Let S be a set. A *d-system* on S is a family $\mathcal{D}$ of subsets of S such that: 1) m $S \in \mathcal{D}$; 2) If $A, B \in \mathcal{D}$ and $A \subset B$, then $B \setminus A \in \mathcal{D}$; 3) If $(A_n)_{n \in \mathbb{N}}$ is an increasing sequence in $\mathcal{D}$, then $\cup_{n \in \mathbb{N}} \in \mathcal{D}$.

Given probabilities μ_i and $i = 1, 2$ on the measurable space $(S, \mathcal{F})$, the family $\mathcal{D} = \{A \in \mathcal{F} \mid \mu_1(A) = \mu_2(A)\}$ in a d-system.

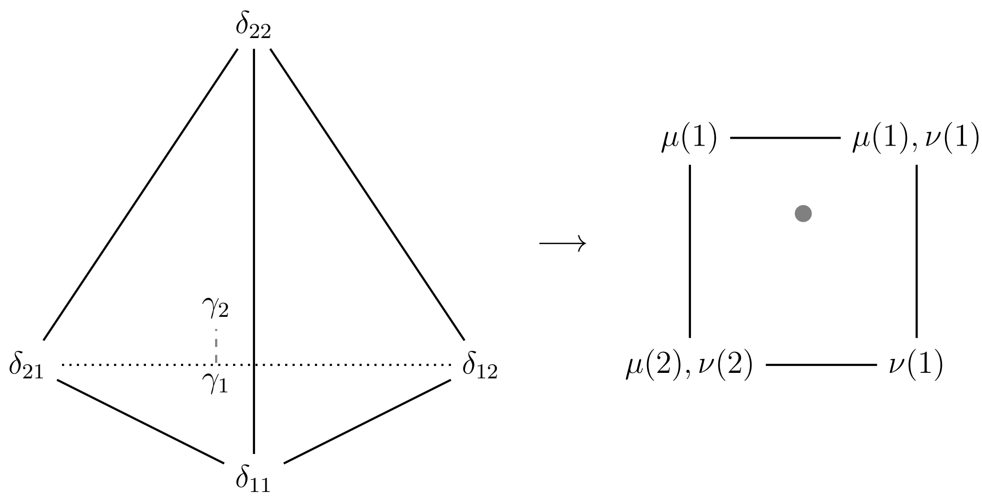

FIGURE 1. 2-dimensional projection of the probability simplex $\Delta(\{1,2\}^2)$. The arrow $\longrightarrow$ represents the marginalization operator with values in $\Delta(\{1,2\}) \times \Delta(\{1,2\})$. The dashed segment from γ_1 to γ_2 represents the coupling polytope.

Given measurable spaces $(S_i, \mathcal{F}_i)$, $i = 1, 2$, the product space $(S, \mathcal{F}) = (S_1 \times S_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma \{A_1 \times A_2 \mid A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$, and $x \in S_1$, the family

$$\mathcal{D} = \{A \in \mathcal{F}_1 \otimes \mathcal{F}_2 \mid A \cap \{x\} \times S_2 = \{x\} \times A_x, A_x \in \mathcal{F}_2\}$$

is a d-system.

A family of subsets of S is a σ-field if, and only if, it is both a d-system and a π-system.

If $\mathcal{I}$ is a π-system, then $d(\mathcal{I}) = \sigma(\mathcal{I})$.

Any d-system that contains a π-system contains the σ-field generated by the π-system.

If two probability measures on the same measurable space agree on a π-system $\mathcal{I}$ they are equal on $\sigma(\mathcal{I})$.

6. Bivariate probability functions with fixed marginals

Consider a finite sample space S and a joint probability function γ. Let μ, ν be the margins, that is the images under the two projection operators, $\Pi_1 \colon S^2 \ni (x, y) \mapsto x$ and $\Pi_2 \colon S^2 \ni (x, y) \mapsto y \in S$. If we present γ as a matrix, $\Gamma = [\gamma(x, y)]_{x,y \in S}$, and probability functions as row vectors, then $\mu = \mathbf{1}^t \Gamma^t$, $\nu = \mathbf{1}^t \Gamma$. Conversely, given the margins (μ, ν) let $\mathcal{P}(\mu, \nu)$ be the set of joint probability functions with the given margins. It is a non-empty, closed, and convex, subset of the probability simples on $S \times S$. Consider the case $S = \{1, 2\}$. See the figure 1.

Let d be a distance on S and define $D(\mu, \nu) = \inf \left\{ \sum_{x,y} d(x, y) \gamma(x, y) \,\middle|\, \gamma \in \mathcal{P}(\mu, \nu) \right\}$. Show that the minimum is reached and that D is a distance on $\Delta(S)$. Compute the distance in the special case above. It is the so-called Kantorovich distance, see `https://en.wikipedia.org/wiki/Wasserstein_metric`.

7. Entropy

The entropy function $H(\lambda) = -\sum_\omega \lambda(\omega)\log\lambda(\omega) = \mathbb{E}_\lambda\left[\phi\circ\lambda\right]$ is defined on the convex set $\Delta^\circ(\Omega)$ of strictly positive probability functions. Set $N = \#\Omega$. The entropy of the uniform probability is $-\sum \frac{1}{N}\log\frac{1}{N} = \log N$. The entropy of a generc λ can be written as

$$H(\lambda) = -\sum_\omega \frac{1}{N}(N\lambda(\omega))\left(\log(N\lambda(\omega)) + \log\frac{1}{N}\right) = \sum_\omega \frac{1}{N}\phi\left(N\lambda(\omega)\right) + H(1/N)\ .$$

As the function $x \mapsto \phi(t) = -t\log t$ is concave, then $\phi(t) - \phi(1) \leq \phi'(1)(t-1)$, that is, $\phi(t) - \phi(1) \leq 1 - t$, hence $H(\lambda) \leq H(1/N)$. Or, Jensen inequality for the concave function *log* gives

$$H(\lambda) = \sum_\omega \lambda(\omega)\log(1/\lambda(\omega)) \leq \log\left(\sum_\omega \lambda(\omega)\frac{1}{\lambda(\omega)}\right) = \log N = H(1/N)\ .$$

The uniform probability function is the unique maximum of the entropy. uiquess follows from strict concavity of from the following argument. Assume there is a $\bar\lambda$ which is a maximum for the entropy and let $\theta \mapsto \lambda(\theta)$ be a differentiable curve in $\Delta^0(\Omega)$ such that $\lambda(0) = \bar\lambda$. Let us compute the derivative and put it equal to zero:

$$\left.\frac{d}{d\theta}H(\lambda(\theta))\right|_{\theta=0} = -\left.\sum_{\omega\in\Omega}(\log\lambda(\omega;\theta) + 1)\lambda'(\omega;\theta)\right|_{\theta=0} = 0\ .$$

The solution is given by the equation $= \sum_{\omega\in\Omega}(\log\lambda(\omega) + 1)\lambda'(\omega;0) = 0$ for a generic curve.

As $\bar\lambda$ is in the $\Delta^0(\Omega)$, for each v in the space parallel to the simplex we can consider the curve $\theta \mapsto \bar\lambda + \theta v$ whose derivative at $\theta = 0$ is v. It follows that for each v we have

$$\sum_{\omega\in\Omega}(\log\bar\lambda(\omega) + 1)v(\omega) = 0$$

hence, $\log\bar\lambda$ is constant that is, $\bar\lambda$ is constant $\bar\lambda(\omega) = 1/\#\Omega$.

8. Positive probability functions

Let S be a finite sample space and $p\colon S \to \mathbb{R}$ a probability function. If $p(x) > 0$, $x \in S$, then $p = \mathrm{e}^v$. Conversely, given a random variable u, there exists a unique constant $\kappa(u)$ such that $p = \mathrm{e}^{u-\kappa(u)}$ is a probability density with respect to some probability function μ. In fact, $\kappa(u) = \mathbb{E}_\mu\left[\mathrm{e}^u\right]$.

If $\mathrm{e}^{u_1-\kappa(u_1)} = \mathrm{e}^{u_2-\kappa(u_2)}$, then $u_1 - u_2 =$ constant. Fix a positive probability function p. Every positive probability function q is of the form $q = \mathrm{e}^{u-\kappa(u)}\cdot p$, with $\mathbb{E}_p\left[u\right] = 0$. There is a 1-to-1 correspondence between q and u. The function $\kappa\colon L^2_0(p) \to]0,+\infty[$ is strictly convex and

$$d\kappa(u)[h] = \mathbb{E}_q\left[h\right]\ ,\quad d^2\kappa(u)[h_1,h_2] = \mathrm{Cov}_q\left(h_1,h_2\right)\ ,$$

where $q = \mathrm{e}^{u-\kappa(u)}\cdot p$.

Consider the case of the Binomial probability fumction.

9. First arrival time

In the Bernoulli model with marginal projections X_j, $j = 1, \dots, n$, show that X_I is independent of X_J, $I \cap J = \emptyset$. Compute the marginal and joint distribution of $S_n = \sum_{j=1}^n X_j$ and $T_n = \inf\{k = 1, \dots, n \mid X_k = 1\}$. Compute all the relevant conditional quantities. [Hint: compute first the joint probability function $\mathbb{P}_\theta(T = k, S_n = h)$, $k = 1, 2, \dots n, +\infty$ and $h = 0, \dots, n-k$ for k finite.]

10. Conditional independence

The properties

$$\begin{aligned} \mathbb{P}(A|B \cap C) &= \mathbb{P}(A|B) \ , \\ \mathbb{P}(C|B \cap A) &= \mathbb{P}(C|B) \ , \\ \mathbb{P}(A \cap C|B) &= \mathbb{P}(A|B)\,\mathbb{P}(C|B) \ , \end{aligned}$$

are equivalent. The property in the first two equations is called *sufficiency* and the property in the last equation is called *conditional independence*.

The Markov property is symmetric in the direction of time. If $X_0, \dots, X_n$ is a MC, then the time-reversed process $Y_h = X_{n-h}$ is a Markov process with transitions

$$\mathbb{P}(Y_{h+1} = x|Y_h = y) = \mathbb{P}(X_{n-h-1} = x|X_{n-h} = y) = \\ \frac{\mathbb{P}(X_{n-h-1} = x, X_{n-h} = y)}{\mathbb{P}(X_{n-h} = y)} = \frac{\mathbb{P}(X_{n-h} = y|X_{n-h-1} = x)\,\mathbb{P}(X_{n-h-1} = x)}{\mathbb{P}(X_{n-h} = y)} = \\ \frac{P_{x,y}\pi_{n-h-i}(x)}{\pi_{n-h}(y)} \ .$$

If moreover the MC is stationary that is $\pi_t = \pi$, then the time-reversed process is a MC with the same invariant distribution and transitions

$$Q_{y,x} = \frac{\pi(x)P_{x,y}}{\pi(y)} \ .$$

Equivalently, we can say that the 2-dimensional distribution are given by

$$\mathbb{P}(X_s = x, X_{s+1} = y) = \pi(x)P_{x,y} = \pi(y)Q_{y,x} \ .$$

A stationary Markov chain is *reversible* if $Q_{y,x} = P_{x,y}$. Equivalently, if π is a probability function such that

$$\pi(x)P_{x,y} = \pi(y)P_{y,x} \ ,$$

we sum the previous relation over x to get

$$\sum_{x \in S} \pi(x)P_{x,y} = \pi(y)\sum_{x \in S} P_{y,x} = \pi(y) \ ,$$

so that π is indeed an invariant probability and the MC constructed from π and P is reversible.

Given a Markov matrix P, if there exists a positive function $\kappa \colon S$ such that $\kappa(x)P_{x,y} = \kappa(y)P_{y,x}$ then we can normalize κ. In such a case we have an immediate way to compute the invariant probability.

11. Random Walk on a graph

Let $G = (S, \mathcal{E})$ be a graph. For each vertex $x \in S$ the *degree* of x, $\deg x$, is the number of edges from x. Let E be the *adjacency matrix* of G. The degree as a row vector is $E\mathbf{1}$. Define the Markov matrix

$$P = \operatorname{diag}(E\mathbf{1})^{-1} E \ .$$

i.e., the transitions

$$P_{x,y} = \begin{cases} \frac{1}{\deg x} & \text{if } y \text{ is connected with } x, \\ 0 & \text{if } y \text{ is not connected with } x. \end{cases}$$

Observe that x is connected to y if, and only if, y is connected to x, hence

$$(\deg x) P_{x,y} = (x \to y) = (y \to x) = (\deg y) P_{y,x} \ .$$

It follows that the invariant probability is

$$\pi(x) = \frac{\deg x}{\sum_{y \in S} \deg y} \ .$$

and the MC is reversible.[1]

References

1. Paul Malliavin, *Integration and probability*, Graduate Texts in Mathematics, vol. 157, Springer-Verlag, 1995, With the collaboration of Héléne Airault, Leslie Kay and Gérard Letac, Edited and translated from the French by Kay, With a foreword by Mark Pinsky. MR MR1335234 (97f:28001a)
2. Sheldon M. Ross, *Introduction to Probability Models*, 10th ed., Academic Press, 2010.

[1] If interested, check `https://en.wikipedia.org/wiki/PageRank`