HOMEWORK 6

654 STOCHASTIC PROCESSES

Note for the current and future homework: due to a large class size and time constraints, not all exercises will be graded.

As you are the first class to go through these homework assignments, they may contain typos/ambiguities. Feel welcome to contact us if you see a typo or have doubt about other issues.

Exercises from the book page

- 4.1
- 4.2
- 4.3
- (Optional) 4.4. Note: if $x_i \in \mathcal{X}_i$ for each $i = 1, ..., n$, then the distributions μ and ν on $\mathcal{X}_1 \times ... \times \mathcal{X}_n$ are defined by $\mu((x_1, ..., x_n)) = \mu_1(x_1)\mu_2(x_2)...\mu_n(x_n)$ and $\nu((x_1, ..., x_n)) = \nu_1(x_1)\nu_2(x_2)...\nu_n(x_n)$.
- (Optional) 4.5. Note: $\|f\|_p$ is defined at the beginning of section 4.7. *Hint: use Jensen's inequality, that is, $g(\mathbf{E}[Y]) \leq \mathbf{E}[g(Y)]$ provided that g is convex.

Coupling

1. Let $0 < a < b$. Show that $\|Poisson(b) - Poisson(a)\|_{TV} \leq 1 - \exp(-(b-a)) \leq b - a$. *(Hint: construct a coupling by generating two independent random variables with distributions $Poisson(a)$ and $Poisson(b-a)$).*
2. Let $\lambda > 0$. Show that $\|Poisson(\lambda) - \mu\|_{TV} \leq \frac{C}{n}$ for some absolute constant $C > 0$ where μ is the distribution $Binomial(n, 1 - e^{-\lambda/n})$. *(Hint: construct a coupling by generating n independent $P_i \sim Poisson(\lambda/n)$ random variables and define n Bernoulli random variables by $B_i = \min(1, P_i)$).*
3. (Optional) Consider the graph with vertices $V = \{(i,j) \in \mathbb{Z}^2 : -n \leq i \leq n, -n \leq j \leq n\}$. Two nodes $x \sim y$ are said to be neighbors if $\|x - y\| = 1$. Independently for each pair of neighbors, we put an undirected edge between the two neighbors with some probability $p \in (0,1)$, and let $\mathbf{P}_p$ be the probability with respect to these random edges. Let E be the event that there exists a path from (0,0) to the boundary of the square (the boundary is the set of points (i,j) such that $\max(|i|, |j|) = n$). Prove that if $0 < p < q < 1$ then
$$\mathbf{P}_p(E) \leq \mathbf{P}_q(E).$$

Hint: you may construct a coupling with many independent uniform $[0, 1]$ *random variables.* For an illustration, check the first picture at http://mysite.du.edu/~yyi2/percolation.html.

Compute the optimal coupling

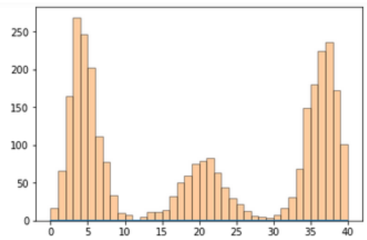
1. Using python or R, generate the optimal coupling (X, Y) of two distributions with domain $\{0, 1, ..., n\}$. The function should take as input two lists (or two arrays), each of length $n+1$, which represent the two pmf. The function should output a random variable (X, Y) distributed according to an optimal coupling.

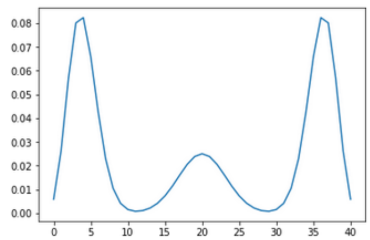
2. Set $n = 40$ and generate many points (X_i, Y_i) as in the previous question when the input distributions are $Binomial(40, 0.45)$ and $Binomial(40, 0.55)$. Plot a heatmap of these points (X_i, Y_i) together with the marginals, using for instance https://seaborn.pydata.org/generated/seaborn.jointplot.html (preferred, to see the empirical marginals on the side) or https://stackoverflow.com/questions/2369492/generate-a-heatmap-in-matplotlib-using-a-scatter-data-set/42568219#42568219. For comparison, generate a similar plot of many independent copies of the coupling (X, Y) of the same two distributions when the two coordinates X and Y are independent.

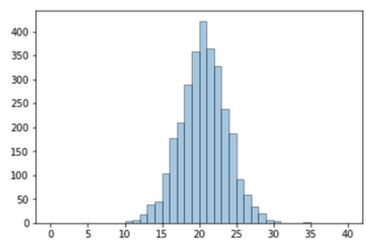
3. Still with $n = 40$, draw new pictures with two distributions of your choice as input.

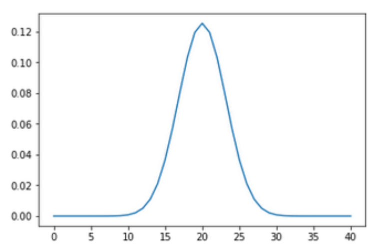
The function `numpy.random` may be useful. See also `scipy.stats.binom.pmf` at https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.stats.binom.html.

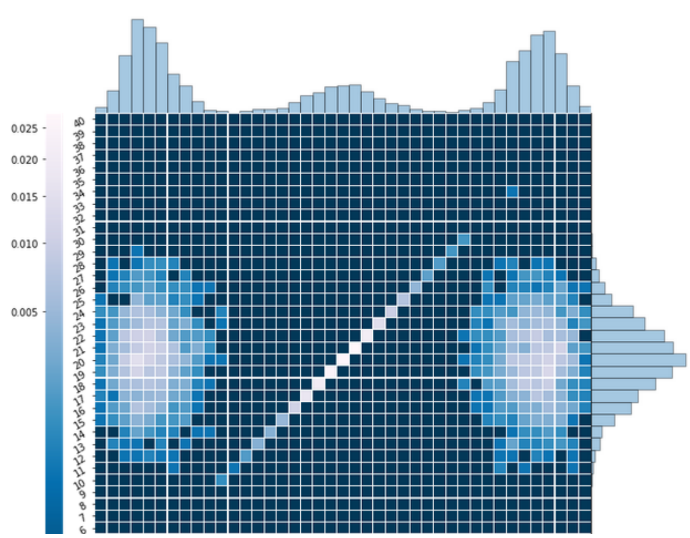
For instance, an optimal coupling will typically look as in the picture.


0
50
100
150
200
250
0 5 10 15 20 25 30 35 40


0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0 5 10 15 20 25 30 35 40


0
50
100
150
200
250
300
350
400
0 5 10 15 20 25 30 35 40


0.00
0.02
0.04
0.06
0.08
0.10
0.12
0 5 10 15 20 25 30 35 40


0.025
0.020
0.015
0.010
0.005

Coupling of Markov Chains (Optional).

Let $\mathcal{X} = \{0, 1, 2, ..., 100\}$ and let P be a transition matrix on $\mathcal{X}$ such that $P(i, j) = e^{-\beta|i-j|}/Z_i$ where Z_i is a normalizing constant and $\beta > 0$ is a parameter that we take as $\beta = 0.1$.

1. Implement a function in R or python that takes as input a distribution μ on $\{0, 1, ..., 100\}$ and outputs a random variable distributed as μ. *Hint: in python, check `numpy.random`.
2. We define a stochastic process $\{(X_t, Y_t), t = 0, 1, 2, ...\}$ as follows.
   - At $t = 0$, define $X_0 = 0$ and $Y_0 = 100$.
   - Given $(X_t, Y_t) = (x_t, y_t)$, generate independently of the previous moves as follows. Generate X_{t+1} according to $P(x_t, \cdot)$,
     - if $x_t = y_t$, defined $Y_{t+1} = X_{t+1}$;
     - If $x_t \neq y_t$, then generate Y_{t+1} according to $P(y_t, \cdot)$ (independently of X_{t+1}).
3. Explain why $\{X_t, t = 0, 1, 2, ...\}$ is a Markov Chain with transition matrix P.
4. Generate a few realizations of the previous process up to time 200. Draw the two chains (X_t) and (Y_t) on the same graph (where the horizontal axis represents the time, and draw a line between X_t and X_{t+1} and similarly for (Y_t)).
5. Generate 100 copies of the process $\{(X_t, Y_t), t = 0, 1, 2, ...\}$ up to time 200 and estimate $\mathbf{E}[\min\{t = 0, ..., 200 : X_t = Y_t\}$ using the law of large numbers with $n = 100$.

When the two chains X_t and Y_t are equal for the first time, we sometimes say that they *coalesce*. We now try to construct a different coupling, where the chains coalesce faster.

a. Implement in python or R a function that
   - takes as input two nodes distributions μ, ν over $\mathcal{X}$.
   - outputs two random variables (X, Y) distributed as the optimal coupling between μ and ν. (cf. the proof of Proposition 4.7 for the definition of the optimal coupling). In python, the functions `numpy.random`, `numpy.abs`, `numpy.minimum` may be useful.

b. Implement in python or R a function that
   - takes as input two nodes x, y
   - outputs two random variables (X, Y) distributed as the optimal coupling between $P(x, \cdot)$ and $P(y, \cdot)$.

c. We define a stochastic process $\{(X_t, Y_t), t = 0, 1, 2, ...\}$ as follows.
   - At $t = 0$, define $X_0 = 0$ and $Y_0 = 100$.
   - Given $(X_t, Y_t) = (x_t, y_t)$, generate (X_{t+1}, Y_{t+1}) independently of the previous moves from the optimal coupling between $P(x_t, \cdot)$ and $P(y_t, \cdot)$.

d. Does $X_t = Y_t$ imply $X_{t+1} = Y_{t+1}$?

e. Explain why $\{X_t, t = 0, 1, 2, ...\}$ is a Markov Chain with transition matrix P.

f. Generate a few realizations of the previous process up to time 200. Draw the two chains (X_t) and (Y_t) on the same graph (where the horizontal axis represents the time, and draw a line between X_t and X_{t+1} and similarly for (Y_t)).

g. Generate 100 copies of the process $\{(X_t, Y_t), t = 0, 1, 2, ...\}$ up to time 200 and estimate $\mathbf{E}[\min\{t = 0, ..., 200 : X_t = Y_t\}$.