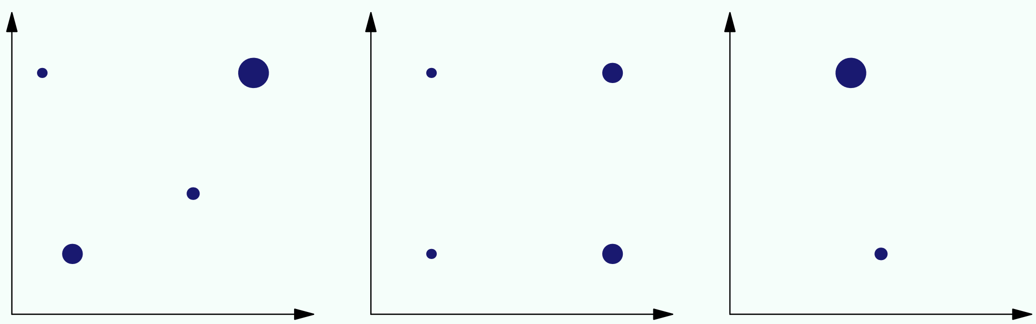
DATA 1010
Problem Set 8
Due 02 November 2018 at 11 PM

Problem 1

Each of the following figures shows the PMF of the joint distribution of two random variables (where point size is reflective of probability mass).



(a) For each figure, indicate whether the two random variables are positively correlated, negatively correlated, or uncorrelated.

(b) In which figure are the random variables independent?

(c) In the second figure, which random variable is uniformly distributed on its support?

(d) Estimate the mean of the random vector $[X, Y]$ for the third distribution (assuming that the two random variables are X and Y). Express your answer by drawing on the figure.

Problem 2

Suppose that X is chosen uniformly at random from the interval $[0, 1]$, and then Y is chosen uniformly at random from the interval $[0, X]$.

(a) The information above specifies $f_{Y \mid X=x}(y)$, the conditional density of Y given that $X = x$. Find $f_{Y \mid X=x}(y)$.

(b) Use (a) to find the joint density $f_{X,Y}(x, y)$ of X and Y.

(c) Use (b) to find the marginal distribution of Y.

(d) You should find that the density of Y is *unbounded*. Explain why it isn't a contradiction for a probability density function to be unbounded (considering that the total amount of probability mass must be 1).

Problem 3

The Student's t-distribution with parameter ν is the distribution of the random variable

$$\frac{\overline{X}_n - \mu}{S_n / \sqrt{n}},$$

where $n = \nu + 1$, where $X_1, \ldots, X_n$ is a sequence of independent $\mathcal{N}(\mu, \sigma^2)$'s, where $\overline{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$, and where $S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2}$.

Estimate the variance of the Student's t-distribution with parameter $\nu = 10$ by using the above description to sample from it M times for some large M. Then compute the variance of the distribution which places probability mass $1/M$ at each of the simulated samples.

Look up the exact formula for the variance of the Student's t-distribution on Wikipedia and check that your result is close to the true value.

Problem 4

Recall the probability mass function struct `PMF` we defined in Problem Set 7. Define a sampling method for it. Your function should return a random value whose distribution is the one represented by the given PMF.

```
function sample(P::PMF)
# ...
end
```

Problem 5

A random walker begins at one vertex of a square, and it repeatedly moves along one of the adjacent edges (chosen uniformly at random) to another vertex. Find the distribution of the number of steps N when the walker first reaches the vertex diagonally opposite to the starting vertex.

Problem 6

In this problem we will show that a sum Z of independent standard normal random variables X and Y is normal with mean zero and variance 2.

(a) Let F_Z be the CDF of Z. Express $F_Z(z)$ as an integral over a subset of $\mathbb{R}^2$.

(b) You should find that the integrand in your answer to (a) is rotationally symmetric. Rotate the region of integration so that its boundary is a vertical line.

(c) Simplify the integral you found in (b) and show that it is equal to $\Phi_{0,2}(z)$, where Φ_{μ,σ^2} denotes the CDF of $\mathcal{N}(\mu, \sigma^2)$.

Note: this idea can be extended to show that an $\mathcal{N}(\mu_1, \sigma_1^2)$ plus an independent $\mathcal{N}(\mu_2, \sigma_2^2)$ has distribution $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Problem 7

A **call option** is a financial contract between two parties which grants the buyer the right, but not the obligation, to purchase a specified security at a specified price (called the **strike price**) at a specified date in the future (called the **expiration date**).

Suppose that you purchase a call option for 10 shares of AAPL with a strike price of \$216 and an expiration 22 business days from now. Suppose that the daily change in the price of AAPL is normally distributed with mean zero and standard deviation \$8.44, and that the changes for different days are independent.

(a) Find a function f such that the call option is worth $f(P)$ dollars to you if the share price in 22 days is P. Draw a graph of f. Hint: if the price is greater than \$216, would you exercise the option and purchase the stock? What if it's less than \$216?

(b) Find the distribution of P.

(c) Find the fair price of the call option, based on the above assumptions.

Notes: (1) the data in this problem are real: the current price at time of writing is \$216, and the daily fluctuations have had an empirical standard deviation of \$8.44 historically. The number of business days in a month is approximately 22. (2) Although this problem uses finance ideas, all of the finance information you need to solve the problem is in the problem statement.

Problem 8

Simulate $n = 1000$ samples from the joint distribution of X and Y, given that X is uniform on $[0, 1]$ and $Y = 2 + 1.2X + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 0.5)$. Record the integrated squared error for the Nadaraya-Watson estimator (with bandwidth selected by cross-validation) and for the line of best fit.

Notes: you can approximate the integrated squared difference between two functions by evaluating the squared difference at the points of a fine-mesh grid.