Caltech Department of Mathematics

Ma 3/103 KC Border
Introduction to Probability and Statistics Winter 2020

Assignment 2: Exercises on Expectation; Bayes' Law

Due Tuesday, January 21 by 4:00 p.m.
in the dropbox in the lobby of Linde Hall
(the building formerly known as Sloan).

Instructions:

When asked for a probability or an expectation, give both a formula and an explanation for why you used that formula, and also give a numerical value when available.

When asked to plot something, use informative labels (even if handwritten), so the TA knows what you are plotting, attach a copy of the plot, and, if appropriate, the commands that produced it.

No collaboration is allowed on optional exercises.

When a problem says that an element of a set is selected **at random**, assume that each element is equally likely to be chosen.

Exercise 1 (30 pts) Let X and Y be independent Bernoulli(1/2) random variables (coin tosses), and let Z be the parity of $X+Y$. (That is, $Z = 1$ if $X+Y$ is odd, and 0 otherwise)

Prove that X, Y, and Z are pairwise stochastically independent (any pair are independent); but X, Y, Z are *not* mutually stochastically independent. □

Exercise 2 (Introduction to Bayesian Inference) (50 pts)

There are n urns filled with black and white balls. Let f_i be the fraction of white balls in urn i. In stage 1 an urn is chosen at random (each urn has probability $1/n$ of being chosen). In stage 2 a ball is drawn at random from the urn. Thus the sample space is $S = \{1, \dots, n\} \times \{B, W\}$. Let every subset of S be an event.

1. (10 pts) Draw a "tree diagram" (see Pitman [1, § 1.6]) describing this experiment.

Let U_i be the event that urn i was selected at stage 1. Let W denote the event that a white ball is drawn at stage 2, let B denote the event that a black ball is drawn at stage 2.

2. (20 pts) Use Bayes' Rule to express $P(U_i|W)$ in terms of $f_1, \dots, f_n$.
3. (10 pts) Use Bayes' Rule to express $P(U_i|B)$ in terms of $f_1, \dots, f_n$.

For concreteness say there are three urns and urn 1 has 30 white and 10 black balls, urn 2 has 20 white and 20 black balls, and urn 3 has 10 white balls and 30 black balls.

4. (10 pts) Compute $P(U_1|W)$, $P(U_1|B)$, $P(U_2|W)$, $P(U_2|B)$, $P(U_3|W)$, and $P(U_3|B)$.

□

Exercise 3 (Problem 3.3.15 in Pitman) (20 pts)

Let X and Y be independent random variables. Show that

$$\boldsymbol{Var}(X - Y) = \boldsymbol{Var}(X + Y).$$

□

Exercise 4 (Cf. problem 3.3.26 in Pitman) (25 pts) Use Jensen's Inequality (Lecture 6) to show that for a random variable X with finite mean μ,

$$\text{std. dev.}\, X \geqslant \boldsymbol{E}\,|X - \mu|,$$

with equality if and only if $|X - \mu|$ is degenerate. □

Exercise 5 (Sen's missing women) (50 pts) For the sake of argument let us assume that the probability of being born a boy $P(B)$ is the same as the probability of being born a girl $P(G)$, namely 1/2. Let us assume that the sex of different children are stochastically independent, and that there are no multiple births, childless families, or adoptions.

In this case we would expect the population to be about half male and half female. Or would we? According to Nobel Prize wining economist Amartya Sen [2], due to differential mortality, in Europe and North America there are about 105 females for every 100 males. But in other countries the ratio is considerably lower. The number of females per 100 males is 94 in China, 93 in India, and 92 in Pakistan, and 84 in Saudi Arabia (which has a large migrant male workforce). These latter countries are sometimes described as having "missing women."

A possible explanation might be that in some countries, parents prefer to have boys so they may continue to have children until they have a boy or maybe two boys. Let's examine the consequences of such behavior.

Let B denote the number of boys, G be the number of girls, and let N denote the total number of children in a family. (These are random variables.) Below are four possible parental decision rules For each rule, write down a parsimonious sample space, and for each representative point in the sample space, show its probability, and the values of each random variable B, G, N, B/N, and G/N at that point. (A table should do nicely.) Compute the expectations $\boldsymbol{E}\,B$, $\boldsymbol{E}\,G$, $\boldsymbol{E}\,N$, $\boldsymbol{E}\,B/\boldsymbol{E}\,N$, $\boldsymbol{E}\,G/\boldsymbol{E}\,N$, $\boldsymbol{E}(B/N)$, and $\boldsymbol{E}(G/N)$. Here are the four rules to consider:

1. Parents have exactly one child.
2. Parents stop having children once they have a boy or two girls, whichever comes first.
3. Parents always have two children.
4. Parents have children until they have a boy. Take the idealization that the family has no limit on the number of children. (My own great-great-grandfather had 22 children that survived infancy. Not all of his three wives survived childbirth.)

   This sample space will have infinitely many points, so you can't list them all, but you can give a description of the n^{th} point. Then each expectation will be given by an infinite series. Show both the series and its value.

(Note: I had a hard time with some of the infinite series involved here, but you may not. But if you do, you may want to consult the supplementary notes. "Some sums," on the auxiliary course web site.)

□

Exercise 6 (10 pts) How much time did you spend on the previous exercises? □

Exercise 7 (Optional Exercise) (50 pts) There are n balls numbered $1, \dots, n$ and n bins numbered $1, \dots, n$. The balls are put into the bins at random, one per bin. For each $k = 0, \dots, n$, what is the probability that exactly k balls are put in the matching bin? Explain your reasoning. □

References

[1] J. Pitman. 1993. *Probability*. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.

[2] A. K. Sen. 1998. Mortality as an indicator of economic success and failure. *Economic Journal* 108(446):1–25. http://www.jstor.org/stable/2565734