Dr. Z.'s Probability Lecture 17 Handout: Expectation of Sums of Random Variables

By Doron Zeilberger

Very Important Fact (Discrete Case): If X and Y have a joint probability mass function $p(x,y)$ then

$$E[\,g(X,Y)\,] \,=\, \sum_y \sum_x g(x,y)p(x,y) \quad .$$

Problem 17.1: In a certain community every family has at most ten dogs and at most ten cats. Let X be the number of dogs and Y the number of cats. It turns out that the joint probability mass function is

$$p(x,y) = \begin{cases} \frac{2x+3y}{3025} \quad if \quad 0 \leq x \leq 10 \quad and \quad 0 \leq y \leq 10\,; \\ 0\,, otherwise. \end{cases} \quad .$$

(i) Find $E[XY]$ the expected number of the product of the number of dogs and the number of cats.

(ii) Find $E[|X-Y|]$ the expected number of the difference between the number of dogs and cats. (You can use Maple, in a test, just leave the Maple code to get the number).

Sol. to 17.1(i): Using the important fact

$$E[XY] = \frac{1}{3025}\sum_{i=0}^{10}\sum_{j=0}^{10}(2i+3j)ij \,=\, \frac{1}{3025}\sum_{i=0}^{10}\sum_{j=0}^{10}(2i^2j+3ij^2) \,=\, \frac{1}{3025}\sum_{i=0}^{10}\sum_{j=0}^{10}(2i^2j+3ij^2) \,=$$

$$\frac{1}{3025}\sum_{i=0}^{10}\sum_{j=0}^{10}2i^2j \,+\, \frac{1}{3025}\sum_{i=0}^{10}\sum_{j=0}^{10}3ij^2 \,=$$

$$\frac{2}{3025}\sum_{i=0}^{10}\sum_{j=0}^{10}i^2j \,+\, \frac{3}{3025}\sum_{i=0}^{10}\sum_{j=0}^{10}ij^2 \quad .$$

By symmetry, the two double-sums are the same, so this equals

$$\frac{5}{3025}\sum_{i=0}^{10}\sum_{j=0}^{10}i^2j \,=\, \frac{1}{605}\left(\sum_{i=0}^{10}i^2\right)\left(\sum_{j=0}^{10}j\right) \,=\, \frac{1}{605}\,\frac{10\cdot 11\cdot 21}{6}\cdot\frac{10\cdot 11}{2} \,=\, 35.$$

(Here we used the formulas $\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$, and $\sum_{j=1}^n j = \frac{n(n+1)}{2}$).

Note: If you have Maple, just type

`add(add((2*i+3*j)*i*j/3025,i=0..10),j=0..10);` and get 35.

Ans. to 17.1(i): $E[XY] = 35$.

Sol. to 17.1(ii):

Using Maple, this equals `add(add((2*i+3*j)*abs(i-j)/3025,i=0..10),j=0..10);` and Maple says that it is $\frac{40}{11}$

Ans. to 17.1(ii): $E[|X-Y|] = \frac{40}{11}$.

Very Important Fact (Continuous Case): If X and Y have a joint density function $f(x,y)$ then

$$E[\, g(X,Y)\,] \;=\; \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y)\, dx\, dy \quad .$$

Note: When $f(x,y)$ is non-zero in a finite region, the limit of integration would be finite numbers, according to the definition of $f(x,y)$.

Problem 17.2: An accident occurs at a point X, along a highway L miles long, that has a density function

$$a(x) \;=\; \begin{cases} \frac{2x}{L^2}\,,\ if \quad 0 \le x \le L\,; \\ 0\,,\ otherwise. \end{cases} \quad .$$

At the time of the accident, an ambulance is at location Y that has a density function

$$b(y) \;=\; \begin{cases} \frac{3y^2}{L^3}\,,\ if \quad 0 \le y \le L\,; \\ 0\,,\ otherwise. \end{cases} \quad .$$

Assuming that the location of the accident and the ambulance are **independent**, find the expected distance between the ambulance and the accident. Leave your answer as a sum of double integrals that Maple can do.

Sol. of 17.2: By independence, the joint density function is

$$f(x,y) \;=\; \begin{cases} \frac{6xy^2}{L^5}\,,\ if \quad 0 \le x,y \le L\,; \\ 0\,,\ otherwise. \end{cases} \quad .$$

Hence

$$E[|X-Y|] = \frac{6}{L^5} \int_0^L \int_0^L xy^2 |x-y|\, dx\, dy \quad .$$

$$\frac{6}{L^5} \int_0^L \int_0^y xy^2 (y-x)\, dx\, dy + \frac{6}{L^5} \int_0^L \int_y^L xy^2 (x-y)\, dx\, dy \quad .$$

The Maple command is

```
6/L**5* int(int(x*y**2*(y-x),x=0..y),y=0..L)+6/L**5* int(int(x*y**2*(x-y),x=y..L),y=0..L);
.
```

and you get $L/4$.

Ans. to 17.2: $E[|X-Y|] = \frac{L}{4}$.

Trivial But VERY useful fact (Linearity of Expectation): If $X_1, X_2, \ldots X_n$ are *any* random variables (on the same probability space), then

$$E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n] \quad .$$

Some Applications

Boole's Inequality

For any events

$$P\left(\bigcup A_i\right) \leq \sum_{i=1}^{n} P(A_i) \quad .$$

Expectation of a binomial random variable

Fact: When you toss a coin, whose probability of Heads is p, n times, the expected number of Heads is np.

Proof: Let X be the random variable "total number of Heads". Let X_i be the random variable "number of Heads at the i-th toss". Then $X = X_1 + \ldots + X_n$.

But

$$E[X_i] = p \cdot 1 + (1-p) \cdot 0 \, = \, p \quad .$$

Hence

$$E[X] = \sum_{i=1}^{n} E[X_i] = np \quad .$$

Expectation of a negative binomial random variable

Recall that if we repeat a trial whose probability of success is p, the number of times until we score a success is called the *geometric distribution with parameter* p, and its expectation is $\frac{1}{p}$. More generally, the random variable that tells you how long it takes until your score r successes is called the negative binomial distribution with parameters (r, p). Let X_1 be the random variable that tells you the number of tries until the first success, X_2 the random variable that tells you how long it takes between the first success and the second success, etc. Then $X = X_1 + X_2 + \ldots + X_r$. Each of the X_i is a geometric random variable with parameter p, hence $E[X_i] = \frac{1}{p}$, **for every** i, so

$$E[X_1 + \ldots + X_r] = E[X_1] + \ldots + E[X_r] = \frac{r}{p} \quad .$$

The Expected Number of Fixed Points of a Permutation

Let X be the random variable "number of fixed points of a permutation", defined on the set of permutations of length n, namely:

$$X(\pi) = \#\{i \quad ; \quad \pi_i = i\} \quad .$$

For example when $n = 6$, $X(321465) = 2$, since 2 is in the second place and 4 is in the 4-th place.

Very interesting Fact: $E[X] = 1$.

In other words the expected number of fixed points is **always** 1, regardless of n.

Proof: Let X_i be the random variable (that either equals 0 or 1)

$$X_i(\pi) = \begin{cases} 1\,, & if \quad \pi_i = i\,; \\ 0\,, otherwise. \end{cases} \quad .$$

Note that $P(\pi_i = i) = \frac{1}{n}$, since the i-th entry of π is equally likely to be any member of the set $\{1, \ldots, n\}$. Hence

$$E[X_i] = \frac{1}{n} \cdot 1 + \frac{n-1}{n} \cdot 0 \; = \; \frac{1}{n} \quad .$$

Hence

$$E[X] = E[X_1 + \ldots + X_n] = E[X_1] + \ldots + E[X_n] = n \cdot \frac{1}{n} \; = \; 1 \quad .$$

A more complicated example: Let X be the number of fixed points of a permutation, as above, find $E[X^2]$.

Sol.

$$X^2 = \left(\sum_{i=1}^{n} X_i \right)^2 = \sum_{i=1}^{n} X_i^2 + 2 \sum_{1 \le i < j \le n} X_i X_j$$

Since $X_i^2 = X_i$ (why?), $E[X_i^2] = \frac{1}{n}$. Also we have $X_i X_j = 1$ if and only of $\pi_i = i$ and $\pi_j = j$, and the probability of that happening is $\frac{1}{n} \cdot \frac{1}{n-1}$ (why?), hence

$$E[X^2] \; = \; n \cdot \frac{1}{n} + 2 \frac{n(n-1)}{2} \cdot \frac{1}{n(n-1)} = 1 + 1 = 2 \quad .$$

It follows that the variance of X, $Var(X)$, equals $2 - 1^2 = 1$.

The famous coupon collecting problem: Suppose there are N types of coupons and each time one obtains a coupon it is equally likely to be any of the N types. What is the expected number of coupons one needs to collect before obtaining a full collection.

Ans. Let X_i be the number of coupons needed to be purchased in order to get a brand-new coupon, after one already has i coupons. Of course $X_0 = 1$. Once you have i coupons, there are $N - i$ coupons that you don't yet have, so the probability of a coupon that you buy being brand new is $\frac{N-i}{N}$. This is a geometric random variable with parameter $\frac{N-i}{N}$ hence $E[X_i] = \frac{N}{N-i}$. Hence

$$E[X_0 + \ldots X_{N-1}] = \frac{N}{N} + \frac{N}{N-1} + \ldots + \frac{N}{1} = N[1 + \frac{1}{2} + \ldots + \frac{1}{N}] \quad .$$

Problem 17.3: The return on two investments, X and Y, follows the joint probability density function

$$f(x,y) = \begin{cases} 1/2 & , \quad 0 < |x| + |y| < 1 \\ 0 & , \quad otherwise. \end{cases}$$

Calculate (i) $E[2X+Y]$ and (ii) $Var(2X+Y)$.

Set-up the integrals, but do not evaluate them.

Sol. of 17.3: The region where $f(x,y)$ is non-zero is the square whose vertices are $(-1,0),(0,1),(1,0),(0,-1)$ i.e. the region in the plane that stretches on the x axis from $x=-1$ to $x=1$ and

- for $-1 \le x \le 0$, y goes from $y = -1-x$ to $y = 1+x$
- for $0 \le x \le 1$, y goes from $y = -1+x$ to $y = 1-x$.

So the set-up of the integral is

$$\int_{-1}^{0} \int_{-1-x}^{1+x} Whatever\, dy\, dx + \int_{0}^{1} \int_{-1+x}^{1-x} Whatever\, dy\, dx \quad .$$

We have

$$E[2X+Y] = \int_{-1}^{0} \int_{-1-x}^{1+x} \frac{2x+y}{2}\, dy\, dx \, + \int_{0}^{1} \int_{-1+x}^{1-x} \frac{2x+y}{2}\, dy\, dx \quad .$$

That Maple says is 0 (you can also deduce it by symmetry). So the answer to part (i) is 0. Since $Var(2X+Y) = E[(2X+Y)^2] - E[2X+Y]^2$, we need

$$E[(2X+Y)^2] = \int_{-1}^{0} \int_{-1-x}^{1+x} \frac{(2x+y)^2}{2} + \int_{0}^{1} \int_{-1+x}^{1-x} \frac{(2x+y)^2}{2} \quad .$$

That Maple says is $\frac{5}{6}$. Hence $Var(2X+Y) = \frac{5}{6} - 0^2 = \frac{5}{6}$.

The Maple command is

```
int(int(1/2*(2*x+y)**2,y=-1-x..1+x),x=-1..0)+ int(int(1/2*(2*x+y)**2,y=-1+x..1-x),x=0..1);
```

.

Problem 17.4: Let X denote the proportion of employees at a large firm who will choose to be covered under the firm's medical plan, and let Y denote the proportion who will choose to be covered under both the firm's medical and dental plan.

Suppose that for $0 \le y \le x \le 1$, X and Y have the joint cumulative distribution function

$$F(x,y) = y(x^2 + xy - y^2) \quad .$$

Calculate the expectation and variance of the proportion of employees who will choose to be only covered by the medical plan.

Sol. of 17.4: Watch out! What is given is the joint **cumulative** distribution function **not** the joint density function. Hence the first step is to find $f(x, y)$.

$$f(x,y) \;=\; \frac{\partial^2}{\partial x \partial y} y(x^2 + xy - y^2) = \frac{\partial^2}{\partial x \partial y}(x^2 y + xy^2 - y^3) = 2x + 2y \quad .$$

We need $E[X]$ and $E[X^2]$. First let's do $E[X]$. We have:

$$E[X] = \int_0^1 \int_0^x (2x + 2y)x \, dy \, dx = \int_0^1 \int_0^x (2x^2 + 2xy) \, dy \, dx \quad .$$

The inner integral is

$$\int_0^x (2x^2 + 2xy) \, dy \;=\; (2x^2 y + xy^2)\Big|_{y=0}^{y=x} = 2x^2(x - 0) + x(x^2 - 0^2) = 3x^3 \quad .$$

The outer integral is

$$\int_0^1 3\, x^3 \, dx \;=\; \frac{3}{4} x^4 \Big|_0^1 = \frac{3}{4}(1^4 - 0^4) = \frac{3}{4} \quad .$$

Hence $E[X] = \frac{3}{4}$.

Next we need $E[X^2]$.

$$E[X^2] = \int_0^1 \int_0^x (2x + 2y)x^2 \, dy \, dx \;=\; \int_0^1 \int_0^x (2x^3 + 2x^2 y) \, dy \, dx \quad .$$

The inner integral is

$$\int_0^x (2x^3 + 2x^2 y) \, dy \;=\; (2x^3 y + x^2 y^2)\Big|_{y=0}^{y=x} = 2x^3(x - 0) + x^2(x^2 - 0^2) = 3x^4 \quad .$$

The outer integral is

$$\int_0^1 3\, x^4 \, dx \;=\; \frac{3}{5} x^5 \Big|_0^1 = \frac{3}{5}(1^5 - 0^5) = \frac{3}{5} \quad .$$

Finally, $Var(X) = \frac{3}{5} - (\frac{3}{4})^2 = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}$.

Ans. to 17.4: The expected portion of people who only took the medical plan is $\frac{3}{4}$ and its variance is $\frac{3}{80}$.