PS-II

Recep Altar Çiçeksiz

Abstract

Probabilistic Number Theory Problem Section II.

Exercise 1. Consider a group of n people. Assuming their birthdays are independent of each other and that each 365 days are equally likely, find the expected number of pairs who have the same birthday.

Solution. Since we are trying to find the expected "number" of some variable, it is important to remember method of indicators. Let X be the number of pairs that have the same birthday. We have pairs of people, hence let us define

$$I_{ij} = \begin{cases} 1, & B_i = B_j \\ 0, & B_i \neq B_j, \end{cases}$$

where B_i denotes the birthday of i^{th} person. With this we see that

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{1\leq i<j\leq n} I_{ij}\right] = \sum \mathbb{E}[I_{ij}],$$

by independence. We have in the sum $\frac{n(n-1)}{2}$ terms and $\mathbb{P}[I_{ij} = 1] = \frac{1}{365}$, thus $\mathbb{E}[X] = \frac{n(n-1)}{2.365}$.

Exercise 2. Let $X_1, X_2, ...$ be a sequence of i.i.d. continuous random variables. We say that a record occurs at time n if $X_n > \max(X_1, ..., X_{n-1})$. Let Y_n be the number of records by time n. Find $\mathbb{E}[Y_n]$.

Solution.

$$I_i = \begin{cases} 1, & a\ record\ occurs\ at\ time\ i, \\ 0, & otherwise, \end{cases}$$

we observe that $Y_n = \sum_{i=1}^{n} I_i$. We again have $\mathbb{E}[Y_n] = \mathbb{E}\left[\sum I_i\right] = \sum \mathbb{E}[I_i]$ Hence we need to calculate $\mathbb{E}[I_i] = \mathbb{P}[I_i = 1] = \mathbb{P}[X_i > \max(X_1, ..., X_{i-1})]$.

Since they are independent we see that $\mathbb{P}[X_i > \max(X_1, ..., X_{i-1})] = \mathbb{P}[X_j > \max(X_1, ..., X_{j-1}, X_{j+1}, ..., X_i)]$, and we have $\sum_{j=1}^{i} \mathbb{P}[X_j > \max(X_1, ..., X_{j-1}, X_{j+1}, ..., X_i)] = 1$, hence $\mathbb{P}[X_i > \max(X_1, ..., X_{i-1})] = \frac{1}{i}$.

[Intuitively, we can think of the first $i-1$ random variables as dividing the real line into i pieces and that X_i to be bigger than their maximum, it should land to the rightmost piece, hence we have the probability $\frac{1}{i}$.]

Therefore, we have $\mathbb{E}[Y_n] = \sum_{i=1}^{n} \frac{1}{i} \sim \log n$.

Exercise 3. Let X be a discrete random variable whose range is the non-negative integers. Show that

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} 1 - F_X(n),$$

where $F_X(n) = \mathbb{P}[X \leq n]$.

Solution. Let

$$I_i = \begin{cases} 1, & X > i \\ 0, & otherwise, \end{cases},$$

then we see that $N = \sum_{n=0}^{\infty} I_n$. Now, taking expectations we have

$$\mathbb{E}[X] = \sum \mathbb{E}[I_i] = \sum \mathbb{P}[X > n] = \sum [1 - F_X(n)].$$

Exercise 4. Let $(A)_\beta)_{\beta \in B}$ be a collection of pairwise disjoint events. Show that if $\mathbb{P}(A_\beta) > 0$ for each β, then B must be countable.

Solution. Suppose on the contrary that B is uncountable. Then considering the sets $B_n := \{\beta \in B : \mathbb{P}(A_\beta) > \frac{1}{n}\}$, we observe that $\cup_{n \in \mathbb{N}} B_n = B$. Since the countable union of finite sets is countable, there must be an index $i \in \mathbb{N}$, such that B_i is at least countable. But we see that

$$\mathbb{P}(\cup_{B_i} A_\beta) = \sum_{\beta \in B_i} \mathbb{P}(A_\beta) > \sum_{\beta \in B_i} \frac{1}{i},$$

which diverges. However the whole probability must be 1, thus we reach a contradiction. The set cannot be uncountable.

Exercise 5. Let $\lambda > 0$. Let Y be a random variable whose pmf is given by

$$f(y) = \begin{cases} \frac{e^{-\lambda} \lambda^y}{y!}, & y = 0, 1, 2, ... \\ 0, & otherwise. \end{cases}$$

Find $\mathbb{E}[Y]$.

Solution. We have

$$\sum_{y=0}^{\infty} y\frac{e^{-\lambda}\lambda^y}{y!} = \lambda e^{-\lambda}\sum \frac{\lambda^{y-1}}{(y-1)!} = \lambda.$$

Exercise 6. Let X_1, X_2 be independent Poisson random variables with parameter λ.

(a)What is the distribution of $X_1 + X_2$?

(b)What is the conditional distribution of $X_1 + X_2$ given that $X_1 = k$ with $k \geq 0$?

(c)What is the conditional distribution of X_1 given that $X_1 + X_2 = k$, $k \geq 0$?

Solution. (a) Let us calculate $\mathbb{P}[X_1 + X_2 = k]$. We see that

$$\mathbb{P}[X_1{+}X_2 = k] = \sum_{i=0}^{k}\mathbb{P}[X_1{+}X_2 = k|X_1 = i]\mathbb{P}[X_1 = i] = \sum_{i=0}^{k}\mathbb{P}[X_2 = k{-}i|X_1 = i]\mathbb{P}[X_1 = i].$$

Since they are independent we see that $\mathbb{P}[X{+}2 = k{-}i|X_1 = i] = \frac{\mathbb{P}[X_2=k-i\cap X_1=i]}{\mathbb{P}[X_1=i]} = \frac{\mathbb{P}[X_2=k-i]\mathbb{P}[X_1=i]}{\mathbb{P}[X_1=i]} = \mathbb{P}[X_2 = k - i]$. Hence we have

$$\mathbb{P}[X_2{+}X_1 = k] = \sum_{i=0}^{k}\mathbb{P}[X_2 = k{-}i]\mathbb{P}[X_1 = i] = \sum_{i=0}^{k}\frac{e^{-\lambda}\lambda^{k-i}}{(k-i)!}\frac{e^{-\lambda}\lambda^{i}}{i!} = \frac{e^{-2\lambda}\lambda^k}{k!}\sum_{i=0}^{k}\binom{k}{i} = \frac{e^{-2\lambda}(2\lambda)^k}{k!}.$$

Hence we see that the sum is Poisson r.v. with parameter 2λ.

<u>*Remark*</u>. It is also true for λ and μ. That is $X_1 + X_2$ is a Poisson r.v. with parameter $\lambda + \mu$ if X_1 and X_2 are with parameters λ and μ respectively.

(b)We have

$$\mathbb{P}[X_1 + X_2 = n|X_1 = k] = \mathbb{P}[X_2 = n - k] = \frac{e^{-\lambda}\lambda^{n-k}}{(n-k)!},$$

if $n > k$. Otherwise since both are non-negative the probability is zero.

(c)We have

$$\mathbb{P}[X_1 = n|X_1{+}X_2 = k] = \frac{\mathbb{P}[X_1 = n \cap X_1 + X_2 = k]}{\mathbb{P}[X_1 + X_2 = k]} = \frac{\mathbb{P}[X_1 = n]\mathbb{P}[X_2 = k - n]}{\mathbb{P}[X_1 + X_2 = k]}$$

$$= \frac{\frac{e^{-\lambda}\lambda^n}{n!}\frac{e^{-\lambda}\lambda^{k-n}}{(k-n)!}}{\frac{e^{-2\lambda}(2\lambda)^k}{k!}} = (\frac{1}{2})^k\binom{k}{n}.$$

Hence it turns out to be binomial with paramater $\frac{1}{2}$.

Exercise 7. Five percent of computer parts produced by a certain supplier are defective.

(a)What is the probability that a randomly and independently selected sample of 12 parts contains at least 3 defective ones?

(b)What is the expected number of defective parts among this sample of 12 parts?

Solution. (a)First let us calculate $\mathbb{P}[X = k]$, where X is the number of defective samples among this 12 parts. We can choose this k parts with $\binom{12}{k}$, and the probability that this k parts are defective exactly is $(\frac{1}{20})^k(\frac{19}{20})^{12-k}$. Hence we have

$$\mathbb{P}[X \geq 3] = \sum_{k=3}^{12} \binom{12}{k} (\frac{1}{20})^k (\frac{19}{20})^{12-k}.$$

(b)Let I_k be the indicator function of part k being defective, and Y be the expected number of defective parts. We have

$$\mathbb{E}[Y] = \sum_{i=1}^{1} 2\mathbb{E}[I_k] = \frac{12}{20} = \frac{3}{5}.$$

Exercise 8. Let $X_1, ..., X_n$ be independent random variables with finite mean μ and variance σ^2. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean of $X_1, ..., X_n$.

(i)Find the expectation of $\overline{X}$.

(ii)Find the variance and standard deviation of $\overline{X}$.

(iii)Show that $Cov(X_i - \overline{X}, \overline{X}) = 0$ for each $i = 1, ..., n$.

Solution. (i)

$$\mathbb{E}[\overline{X}] = \frac{1}{n}\sum \mathbb{E}[X_i] = \frac{n\mu}{n} = \mu.$$

(ii)Since they are independent we have

$$Var(\overline{X}) = \frac{1}{n^2}\sum Var(X_i) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

The standard deviation then becomes $\frac{\sigma}{\sqrt{n}}$.

(iii)We have $Cov(X_i - \overline{X}, \overline{X}) = \mathbb{E}[(X_i - \overline{X})\overline{X}] - \mathbb{E}[X_i - \overline{X}]\mathbb{E}[\overline{X}] = \mathbb{E}[X_i\overline{X}] - \mathbb{E}[\overline{X}^2] - \mathbb{E}[X_i]\mathbb{E}[\overline{X}] + \mathbb{E}[\overline{X}]^2$. The second and the fourth term together is $-Var(\overline{X})$, and we know the third term which equals μ^2, hence we have

$$= \mathbb{E}[X_i\overline{X}] - \mu^2 - \frac{\sigma^2}{n}.$$

In order to calculate the first term, we open it:

$$\mathbb{E}[X_i\overline{X}] = \frac{1}{n}\left[\sum_{i\neq j}\mathbb{E}[X_iX_j] + \mathbb{E}[X_i^2]\right] = \frac{\mu^2(n-1)}{n} + \frac{\sigma^2+\mu^2}{n} = \mu^2 + \frac{\sigma^2}{n},$$

hence the covariance is zero.

Exercise 9. The random pair (X, Y) has the distribution

		X		
		1	2	3
	2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
Y	3	$\frac{1}{6}$	0	$\frac{1}{6}$
	4	0	$\frac{1}{3}$	0

(a)Show that X and Y are dependent.

(b)Give a probability table for (U, V) that have the same marginals as (X, Y) but are independent.

Solution. (a)We need to have $f_{X,Y}((x, y)) = f_X(x)f_Y(y)$ to have independence. Let us calculate the marginals, $f_X(1) = f_X(3) = \frac{1}{4}$; $f_X(2) = \frac{1}{2}$, and $f_Y(2) = f_Y(3) = f_Y(4) = \frac{1}{3}$. We see that $f_{X,Y}((2, 4)) = \frac{1}{3} \neq f_X(2)f_Y(4) = \frac{1}{2}\frac{1}{3} = \frac{1}{6}$.

(b)We take the marginals as same, and let $f_{U,V}(u, v) = f_U(u)f_V(v)$, hence we have the table:

		X		
		1	2	3
	2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
Y	3	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
	4	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$