"Creativity is the ability to introduce order into the randomness of nature."

Eric Hoffer

7 Discrete Random Variables II

Functions of random variables

- for a discrete random variable the PMF of $Y = g(X)$ is

$$P_Y(y) = \sum_{x:g(x)=y} P_X(x)$$

- the expected value is

$$E[Y] = \sum_{x \in S_X} g(x) P_X(x)$$

- given an event B, with $P(B) > 0$ the conditional probability mass function of X is

$$P_{X|B}(x) = \begin{cases} \frac{P_X(x)}{P(B)} & \text{, } \textbf{if } x \in B \\ 0 & \text{, otherwise} \end{cases}$$

- the conditional expected value

$$E[X|B] = \sum_{x \in B} x P_{X|B}(x)$$

- the conditional variance is

$$var(X|B) = E[X^2|B] - E^2[X|B]$$

- a random variable X resulting from an experiment with event space $B_1, B_2, \ldots B_m$ has the PMF

$$P_X(x) = \sum_{i=1}^{m} P_{X|B_i}(x) \cdot P(B_i)$$

Numerical characteristics of random variables

- for a discrete random variable

$$X : \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i\in I}$$

the expected valued or mean $E(X)$ or $M(X)$ is defined as

$$\bar{x} = E(X) = M(X) = \sum_{i\in I} p_i x_i$$

- the moments M_k and the central moments m_k of order k are defined by

$$M_k(X) = M(X^k) = \sum_{i\in I} p_i x_i^k,$$

$$m_k(X) = M((X-\bar{x})^k) = \sum_{i\in I} p_i (x_i - \bar{x})^k$$

- the variance is defined as

$$var(X) = \sum_{i\in I} p_i (x_i - \bar{x})^2$$

and the standard deviation is

$$\sigma(X) = \sqrt{var(X)}$$

- the clasical random variables have the following means and variances

$$\begin{aligned}
X \sim Bin(n,p) &\implies E(X) = np \quad \text{and} \quad var(X) = np(1-p)\\
X \sim Ber(p) &\implies E(X) = p \quad \text{and} \quad var(X) = p(1-p)\\
X \sim Geo(p) &\implies E(X) = \frac{1}{p} \quad \text{and} \quad var(X) = \frac{1-p}{p^2}\\
X \sim Po(\lambda) &\implies E(X) = \lambda \quad \text{and} \quad var(X) = \lambda\\
X \sim NB(r,p) &\implies E(X) = \frac{r}{p} \quad \text{and} \quad var(X) = \frac{r(1-p)}{p^2}\\
X \sim Hyp(N,K,n) &\implies E(X) = n\frac{K}{N} \quad \text{and} \quad var(X) = n\frac{K}{N}\left(1-\frac{K}{N}\right)\frac{N-n}{N-1}
\end{aligned}$$

- the following identities hold

$$E(X+aY) = E(X) + aE(Y)$$

$$var(aX+C) = a^2 var(X)$$

where C is the constant random variable.

- if X and Y are independent:

$$E(XY) = E(X)E(Y)$$

$$var(X+aY) = var(X) + a^2 var(Y).$$

Solved problems

Problem 1. *The random variable X is the number of pages in a facsimile transmission. Based on experience, you have a probability model $P_X(x)$ for the number of pages in each fax you send. The phone company offers you a new charging plan for faxes: 0.10 \$ for the first page, 0.09 \$ for the second page, etc., down to 0.06 \$ for the fifth page. For all faxes between 6 and 10 pages, the phone company will charge 0.50 per fax. (It will not accept faxes longer than ten pages.) Find a function $Y = g(X)$ for the charge in cents for sending one fax.*

Solution: We have to search for a function which associates $1 \to 0.10$, $2 \to 0.09$, $3 \to 0.08$, $4 \to 0.07$ and $5 \to 0.06$ and for $n \geq 6$ the corresponding value is 50. There are different strategies, one idea is to try a polynomial function. Thus, a solution is represented by the function

$$g(x) = \begin{cases} 0.11 - x \cdot 0.01 & \text{for} \quad 1 \leq x \leq 5 \\ 0.50 & \text{for} \quad 6 \leq x \leq 10 \end{cases}$$

Problem 2. *You would like a probability model $P_Y(y)$ for your phone bill under the new charging plan. Let us suppose that the number of pages X has the PMF*

$$P_X(x) = \begin{cases} 0.15 & \text{for } x = 1,2,3,4 \\ 0.10 & \text{for } x = 5,6,7,8 \\ 0 & \text{otherwise} \end{cases}$$

Find the PMF and expected value of Y, the charge for a fax.

Solution: The PMF of $Y = g(X)$ is

$$P_Y(y) = \sum_{x:g(x)=y} P_X(x)$$

and the expected value will be

$$E[Y] = \sum_{x \in S_X} g(x) P_X(x)$$

In order to appply the first formula we have to identify the possible values of Y. These values are 0.06, 0.07, 0.08, 0.09, 0.10 and 0.50. Now, for example

$$P_Y(0.06) = \sum_{x:g(x)=0.06} P_X(x)$$

and the equation $0.11 - x \cdot 0.01 = 0.06$ has only the solution $x = 5$ in the set $\{1,2,3,4,5,6,7,8,\ldots\}$, which represents the set S_X of all possible values of X. Hence

$$P_Y(0.06) = \sum_{x:g(x)=0.06} P_X(x) = P_X(5) = 0.10$$

and step by step one finds

$$Y : \begin{pmatrix} 0.06 & 0.07 & 0.08 & 0.09 & 0.10 & 0.50 \\ 0.10 & 0.15 & 0.15 & 0.15 & 0.15 & 0.30 \end{pmatrix}$$

since

$$P_Y(0.50) = \sum_{x:g(x)=0.50} P_X(x) = P_X(6) + P_X(7) + P_X(8) = 0.30$$

Problem 3. *On the Internet, data is transmitted in packets. In a simple model for World Wide Web traffic, the number of packets N needed to transmit a Web page depends on whether the page has graphic images. If the page has images (event I), then N is uniformly distributed between 1 and 50 packets. If the page is just text (event T), then N is uniform between 1 and 5 packets. Assuming a page has images with probability 1/4, find:*

(a) the conditional PMF $P_{N|I}(n)$

(b) the conditional PMF $P_{N|T}(n)$

(c) the PMF $P_N(n)$

(d) conditional PMF $P_{N|N\leq 10}(n)$

(e) conditional expected value $E[N|N \leq 10]$

(f) conditional variance $Var[N|N \leq 10]$

Solution: Denote by $N|I$ and $N|T$ the number o packets needed when the page has images, respectively just text. Both are uniformly distributed but with different numbers of values

$$P_{N|I}(n) = \begin{cases} \frac{1}{50} & \text{for } n = 1, 2, 3, ..., 50 \\ 0 & \text{otherwise} \end{cases}$$

$$P_{N|T}(n) = \begin{cases} \frac{1}{5} & \text{for } n = 1, 2, 3, 4, 5 \\ 0 & \text{otherwise} \end{cases}$$

c) Since I and T form an event space one has the formula

$$P_N(n) = P_{N|I}(n) \cdot P(I) + P_{N|T}(n) \cdot P(T)$$

thus

$$P_N(n) = \begin{cases} \frac{31}{200} & \text{for } n = 1, 2, 3, 4, 5 \\ \frac{1}{200} & \text{for } n = 6, 7, \ldots, 50 \\ 0 & \text{otherwise} \end{cases}$$

d) Given an event B, with $P(B) > 0$ the conditional probability mass function of N is

$$P_{N|B}(x) = \begin{cases} \frac{P_N(n)}{P(B)} & , \text{if } n \in B \\ 0 & , \text{otherwise} \end{cases}$$

The probability of the event $B = \{N \leq 10\}$ is

$$P(N \leq 10) = 5\frac{31}{200} + 5\frac{1}{200} = \frac{4}{5}$$

Hence

$$P_{N|N\leq 10}(n) = \begin{cases} \frac{P_N(n)}{\frac{4}{5}} & , \text{if } n \leq 10 \\ 0 & , \text{otherwise} \end{cases}$$

and eventually

$$P_{N|N\leq 10}(n) = \begin{cases} \frac{5}{4}\frac{31}{200} & , \text{if } 1 \leq n \leq 5 \\ \frac{5}{4}\frac{1}{200} & , \text{if } 6 \leq n \leq 10 \\ 0 & , \text{otherwise} \end{cases}$$

e) the conditional expected value is

$$E[N|N \leq 10] = \sum_{n=1}^{10} n \cdot P_{N|N\leq 10}(n) = 5 \cdot \frac{5}{4}\frac{31}{200} + 5\frac{5}{4}\frac{1}{200} = 1$$

f) for the conditional variance one has the formula

$$var[N|N \leq 10] = \sum_{n=1}^{10} (n - E[N|N \leq 10])^2 \cdot P_{N|N\leq 10}(n)$$

which is the natural formula of the variance.

Remark

More solved problems can be found in the book of Yates and Goodman or in this handout

Proposed problems

Problem 1. *Suppose X, the time in integer minutes you must wait for a bus has the uniform probability mass function:*

$$P_X(x) = \begin{cases} \frac{1}{20} & , x = 1, 2 \ldots, 20, \\ 0 & , \textit{otherwise} \end{cases}$$

Suppose the bus has not arrived by the eighth minute, what is the conditional PMF of your waiting time X ?

Problem 2. *It can take up to four days after you call for service to get your computer repaired. The computer company charges for repairs according to how long you have to wait. The number of days D until the service technician arrives and the service charge C, in dollars, are described by*

$$P_D(d) = \begin{cases} 0.2 & \text{for } d = 1 \\ 0.4 & \text{for } d = 2 \\ 0.3 & \text{for } d = 3 \\ 0.1 & \text{for } d = 4 \\ 0 & \text{otherwise} \end{cases}$$

and

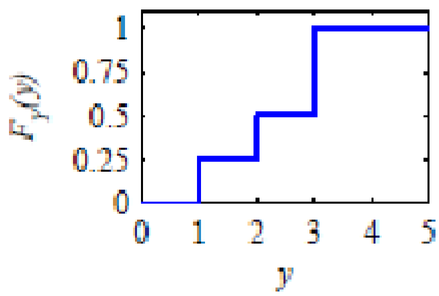
$$C = \begin{cases} 90 & \text{for a 1-day service} \\ 70 & \text{for a 2-day service} \\ 40 & \text{for a 3-day service} \\ 40 & \text{for a 4-day service} \end{cases}$$

a) *What is the expected waiting time $E[D]$?*

b) *Express C as a function of D*

c) *What is the expected value $E[C]$?*

Problem 3. *A discrete random variable Y has the CDF F_Y as shown:*



Use the CDF to find the following probabilities:

a) $P[Y < 1]$

b) $P[Y \leq 1]$

b) $P[Y \geq 2]$

b) $P[Y = 3]$

Problem 4. *The number of memory chips M needed in a personal computer depends on how many application programs A the owner wants to run simultaneously. The number of chips M and the number of application programs A are described by:*

$$M = \begin{cases} 4 & \textbf{\textit{chips for one program}} \\ 4 & \textbf{\textit{chips for two programs}} \\ 6 & \textbf{\textit{chips for three programs}} \\ 8 & \textbf{\textit{chips for four programs}} \end{cases} \quad , \quad P_A(a) = \begin{cases} 0.1(5-a) & a = 1,2,3,4 \\ 0 & \textbf{\textit{otherwise}} \end{cases}$$

a) What is the expected number of programs $\mu_A = E[A]$ *?*

b) Express M*, the number of memory chips, as a function* $M = g(A)$ *of the number of application programs* A*.*

c) Find $E[M] = E[g(A)]$*. Does* $E[M] = g(E[A])$ *?*

Problem 5. *Select some integrated circuits, test them in sequence until you find the first failure, and then stop. Let* N *be the number of tests. All tests are independent with probability of failure* $p = 0.1$*. Consider the condition* $B = \{N \geq 20\}$*.*

i) Find the PMF $P_N(n)$*.*

ii) Find $P_{N|B}(n)$*, the conditional PMF of* N *given that there have been* 20 *consecutive tests without a failure*

iii) What is $E[N|B]$*, the expected number of tests given that there have been* 20 *consecutive tests without a failure?*

Bibliography

[1] R. Yates and D. Goodman. *Probability and Stochastic processes*, Wiley&Sons, 2005.

[2] C. Ariesanu. *Lecture Notes on Special Mathematics*, 2020.