" *Expose yourself to as much randomness as possible.* "

Ben Casnocha

6
Discrete random variables

Texas Holdem Poker

In Hold'em Poker players make the best hand they can combining the two cards in their hand with the 5 cards (community cards) eventually turned up on the table. The deck has 52 and there are 13 of each suit: ♣, ♦, ♥, ♠

Open problem: Until he finally won a WSOP event in 2008, Erick Lindgren was often called one of the greatest players never to have won a WSOP tournament. Before his win, he played in many WSOP events and finished in the top 10 eight times. Suppose you play in one tournament per week. For simplicity, assume that each tournament's results are independent of the others and that you have the same probability p of winning each tournament. If $p = 0.01$, then what is the expected amount of time before you win your first tournament?

Open problem: During Episode 2 of Season 5 of High Stakes Poker, Doyle Brunson was dealt pocket kings twice and pocket jacks once, all within about half an hour. Suppose we consider a high pocket pair to mean 10-10, J-J, Q-Q, K-K, or A-A. Let X be the number of hands you play until you are dealt a high pocket pair for the third time. What is the expected number of hands ?

Open problem: Many casinos award prizes for rare events called jackpot hands. These jackpot hands are defined differently by different casinos. Suppose in a certain casino jackpot hands are defined so that they tend to occur about once every $50,000$ hands on average. If the casino deals about $10,000$ hands per day, what are the expected value and standard deviation of the number of jackpot hands dealt in a 7-day period?

Open problem: On the last hand of the 1998 WSOP Main Event, with the board 8 ♣, 9 ♦, 9 ♥, 8 ♥, 8 ♠, Scotty Nguyen went all-in. While his opponent, Kevin McBride, was thinking, Scotty said

"You call, it's gonna be all over, baby."

McBride said

"I call. I play the board."

It turned out that Scotty had J ♦ , 9 ♣ and won the hand.

Assuming you never fold in the next 100 hands, what would be the expected value of X = the number of times in these 100 hands that you would play the board after all five board cards are dealt ?

Discrete Random Variables

• we'll try to model mathematically different statistical (random) experiments, from the previous handout

Random variables= Mathematical models for random experiments

• the posibile outcomes of such an experiment will be denoted by x_i, $i \in I$, and will be called values of a random variable X

• the probabilities corresponding to each value will form a probability mass function (PMF) denoted $P_X(x)$

• in general, a discrete random variable will be given by its distribution series

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

where $P_X(x_i) = P(X = x_i) = p_i$ means

the probability of the value(outcome) x_i is p_i

• since all the posible outcomes are displayed in a random variable X

$\Longrightarrow \quad \sum_{i \in I} p_i = 1 \quad$ (because 1 means 100%)

Bernoulli's random variable $X \sim Ber(p)$

• it is the simplest discrete random variable

• it models an experiment in which only two possible outcomes can occur, often designated success, and failure.

Example:

It can be used to represent a coin toss. We consider the appearence of a tail being a succes. We assign the value 1 to success with the probability $p \in (0,1)$and the value 0 to failure with probability $q = 1 - p$. Thus we obtain a Bernoulli random variable $X \sim Ber(p)$ with parameter p, the probability of a success

$$X : \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

Of course, in our example $p = \frac{1}{2}$ and

$$P_X(k) = \begin{cases} 1-p, & \text{if } k = 0 \\ p, & \text{if } k = 1 \end{cases}$$

□

Uniform discrete random variable $X \sim \mathcal{U}(n)$

• it represents a mathematical model which generalizes the experiment of throwing a die (case $n = 6$)

• if an experiment has n equally possibile outcomes denoted $\{1, 2, \ldots, n\}$, then the experiment can be modelled using a uniform random variable of the form

$$X : \begin{pmatrix} 1 & 2 & \ldots & n \\ \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \end{pmatrix}$$

• the general form of such a random variable starts with the value k and ends with ℓ, thus it has $\ell - k + 1$ possible values, denote $X \sim \mathcal{U}(k, \ell)$

$$X : \begin{pmatrix} k & k+1 & \ldots & \ell \\ \frac{1}{\ell-k+1} & \frac{1}{\ell-k+1} & \ldots & \frac{1}{\ell-k+1} \end{pmatrix}$$

with the obvious PMF

$$P_X(x) = \begin{cases} \frac{1}{\ell-k+1}, & \text{if } x = k, k+1, \ldots, l \\ 0, & \text{else} \end{cases}$$

Binomial random variable $X \sim Bin(n, p)$

• a random variable with a binomial distribution is the right model when the following assumptions hold (we have a binomial experiment):

• the modelled phemenon consists of n independent trials of the same experiment

• there are only two possible outcomes at each trial (success - failure)

• the probability p of a success is the same at each trial

The random variable which counts the number of successes in n trials of a binomial experiment is called a binomial random variable

$$X : \begin{pmatrix} 0 & 1 & \ldots & k & \ldots & n \\ q^n & C_n^1 p q^{n-1} & \ldots & C_n^k p^k q^{n-k} & \ldots & p^n \end{pmatrix}$$

where p and $q = 1 - p$ are the probabilities of a success, respectively failure at each independent trial.

• thus

$$P_X(k) = \begin{cases} C_n^k p^k q^{n-k}, & \text{for } k \in \{0, 1, 2, \ldots, n\} \\ 0, & \text{else} \end{cases}$$

Geometric random variable $X \sim Geo(p)$

• it is the appropriate model when, in a binomial experiment, we count the number of failures until the first success occurs.

$$X : \begin{pmatrix} 0 & 1 & \dots & k & \dots \\ p & p(1-p) & \dots & p(1-p)^k & \dots \end{pmatrix}$$

• one can easily see

$$P_X(k) = \begin{cases} p(1-p)^k, & \text{for } k \in \{0,1,2,\dots,n\} \\ 0, & \text{else} \end{cases}$$

Hypergeometric random variable $X \sim Hip(N, M, n)$

• consider the problem of drawing objects from a box which contains N objects, with M of them being defective.

• if the draws are with replacement (the extracted object is put back in the box before the next draw), then the number of the defective objects drawn in n draws is a binomial random variable with parameters n and $p = \frac{M}{N}$ (probability of drawing a defective object)

• if the draws are without replacement, then the probability to draw a defective object is not the same in each of those n draws, thus the number of the defective objects drawn is no more a binomial random variable.

$\Longrightarrow$ one obtains a random variable having the probability mass function

$$P(X = k) = \begin{cases} \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, & \text{if } \ k \in \{0,1,2,\dots,n\} \\ 0, & \text{otherwise} \end{cases}$$

and it is called a hipergeometric random variable with parameters M, N and n.

Poisson's random variable $X \sim Po(\lambda)$

• first of all, according to Poisson's law:

$$P(X = k) = C_n^k p^k q^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}, \qquad \text{for } \lambda = np$$

we can approximate the distribution series of a binomial distribution when the probability p of a success at each trial is small and the number of trials n is big. In practice we usually apply it for $p < 0,05$ and $n \geq 100$

• this law generates a probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant mean rate λ and independently of the time since the last event. The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area or volume.

• the Poisson distribution is usually used for rare events and it is also called the law of rare events.

$$X : \begin{pmatrix} 0 & 1 & \dots & k & \dots \\ e^{-\lambda} & \frac{\lambda}{1!}e^{-\lambda} & \dots & \frac{\lambda^k}{k!}e^{-\lambda} & \dots \end{pmatrix}$$

• the PMF of a Poisson random variable is

$$P_X(k) = \begin{cases} \frac{\lambda^k}{k!}e^{-\lambda}, & \text{for } k \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Negative binomial random variable $X \sim NB(r, p)$

• is a discrete probability distribution of the number of successes, in a binomial experiment, before a specified number of failures, denoted r, occurs.

$$X : \begin{pmatrix} 0 & 1 & \dots & k & \dots \\ (1-p)^r & C^1_{1+r-1}p(1-p)^r & \dots & C^k_{k+r-1}p^k(1-p)^r & \dots \end{pmatrix}$$

• sometimes we want to count the number of trials needed to produce the r-th success

• such a random variable X will have the probability mass function:

$$P_X(k) = P(X = k) = C^{r-1}_{k-1}p^r(1-p)^{k-r}, \qquad k = r, r+1, r = 2, \dots$$

The above identity is interpreted as

" the probability to obtain in the k-th trail the r-th success is..."

The expected value $E(X)$ **and the variance** $var(X)$

• the expected value provides a measure of the location or central tendency of a random variable

$$E(X) = \sum_{i \in I} x_i \cdot p_i$$

• the variance $var(X)$ (measure of spread) determines the degree to which the values of a random variable differ from the expected value

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

• as you can see the square distances from every possible value to the expected value are added proportionally to their probability

Solved problems

Problem 1. *Three shooters shoot a target. The random variable X which counts the number of hits has the distribution series*

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{p^2}{4} & \frac{11p}{24} & \frac{1}{4} & \frac{1}{24} \end{pmatrix}.$$

a) After one finds the value of p, compute the probability that X takes a value smaller or equal with 2.
b) Find the probability of hitting the target for each shooter.

Solution: a) The sum of all probabilities in a distribution seris of a random varibale must be 1, thus

$$\frac{p^2}{4} + \frac{11p}{24} + \frac{1}{4} + \frac{1}{24} = 1 \Leftrightarrow 6p^2 + 11p - 17 = 0 \Rightarrow p = 1$$

$$P(X \le 2) = 1 - P(X = 3) = 1 - P(X > 2) = 1 - \frac{1}{24} = \frac{23}{24}$$

b) Let p_1, p_2, p_3 be these probabilities. Hence we have for $p = 1$

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{4} & \frac{11}{24} & \frac{1}{4} & \frac{1}{24} \end{pmatrix}$$

But

$$\frac{1}{4} = P(X = 0) = (1 - p_1)(1 - p_2)(1 - p_3)$$

(because $X = 0$ means: all the shooters missed the target)

$$= 1 - (p_1 + p_2 + p_3) + p_1p_2 + p_1p_3 + p_2p_3 - p_1p_2p_3$$

$$\frac{11}{24} = P(X = 1) = p_1(1 - p_2)(1 - p_3) + p_2(1 - p_1)(1 - p_3) + p_3(1 - p_1)(1 - p_2)$$

(because $X = 1$ means: one of the shooter hit and the others missed the target)

$$= p_1 + p_2 + p_3 - 2(p_1p_2 + p_1p_3 + p_2p_3) + 3p_1p_2p_3$$

$$\frac{1}{4} = P(X = 2) = p_1p_2(1 - p_3) + p_1p_3(1 - p_2) + p_2p_3(1 - p_1)$$
$$= p_1p_2 + p_1p_3 + p_2p_3 - 3p_1p_2p_3$$

$$\frac{1}{24} = P(X = 3) = p_1p_2p_3.$$

One gets the linear system

$$\begin{cases} p_1 + p_2 + p_3 = \frac{13}{12} \\ p_1p_2 + p_1p_3 + p_2p_3 = \frac{3}{8} \\ p_1p_2p_3 = \frac{1}{24} \end{cases}$$

which leads to the equation

$$24x^3 - 26x^2 + 9x - 1 = 0$$

with the roots

$$p_1 = \frac{1}{2}, \quad p_2 = \frac{1}{3}, \quad p_3 = \frac{1}{4}.$$

Problem 2. *When someone presses SEND on a cellular phone, the phone attempts to set up a call by transmitting a SETUP message to a nearby base station. The phone waits for a response and if none arrives within 0.5 seconds it tries again. If it doesn't get a response after $n = 6$ tries the phone stops transmitting messages and generates a busy signal.*

a) Draw a tree diagram that describes the call setup procedure.

b) If all transmissions are independent and the probability is p that a SETUP message will get through, what is the PMF of X, the number of messages transmitted in a call attempt?

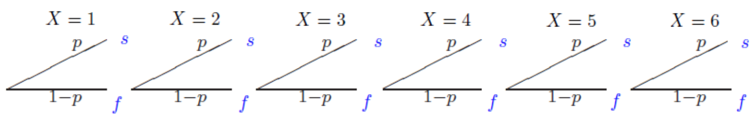
c) What is the probability that the phone will generate a busy signal?

d) As manager of a cellular phone system, you want the probability of a busy signal to be less than 0.02. If $p = 0.9$, what is the minimum value of n necessary to achieve your goal?

Solution:

a) In the setup of a mobile call, the phone will send the SETUP message up to six times. Of course, the phone stops trying as soon as there is a success. Thus we have a geometric random variable $X \sim Geo(p)$ with success probability p. The first value is considered to be $x = 1$.

Using s to denote a successful response, and f a non-response, the sample tree is



b) For this geometric random variable the distribution series will be

$$X : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ p & p(1-p) & p(1-p)^2 & p(1-p)^3 & p(1-p)^4 & (1-p)^6 \end{pmatrix}$$

with the PMF

$$P_X(k) = \begin{cases} p(1-p)^{k-1}, & \text{for } k \in \{1, 2, \ldots, 5\} \\ (1-p)^6, & \text{for } k = 6 \\ 0, & \text{else} \end{cases}$$

c) Let B denote the event that a busy signal is given after six failed setup attempts. The probability of six consecutive failures is $P(B) = (1-p)^6$.

d) To be sure that $P(B) \leq 0.02$, we need to impose the restriction

$$p \geq 1 - (0.02)^{\frac{1}{6}} \approx 48\%$$

Problem 3. *There are 3 traffic barriers along a street. The probability that a car which drives along that street finds any of these three barriers open is $p = 0,8$. We suppose that any of these barriers work independently. Compute:*
a) The distribution series of the random variable which counts the number of barriers passed until the first closed barrier met.
b) Find its cummulative distributionn function.
c) Which is the expected number of barriers found open before the car has to stop in front of a closed one?

Solution: a) We denote by X the desired random variable, which has a distribution series:

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 \\ p_0 & p_1 & p_2 & p_3 \end{pmatrix},$$

where $p_k = P(X = k)$, $k = 0, 1, 2, 3$. By the way we defined the random variable one gets easily:

$$p_0 = P(X = 0) = 0,2$$

$$p_1 = P(X = 1) = 0,8 \cdot 0,2 = 0,16$$

$$p_2 = P(X = 2) = 0,8 \cdot 0,8 \cdot 0,2 = 0,128$$

$$p_3 = P(X = 3) = 0,8 \cdot 0,8 \cdot 0,8 = 0,512$$

Hence:

$$X = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0,2 & 0,16 & 0,128 & 0,512 \end{pmatrix}.$$

b) When $x < 0$ we get by its very definition $F(x) := P(X \leq x) = 0$ because in the interval $(-\infty, 0)$ there are no values of X.

When $0 \leq x < 1$ one gets:

$$F(x) = P(X \leq x) = P(X = 0) = 0,2.$$

When $1 \leq x < 2$ one gets:

$$\begin{aligned} F(x) &= P(X \leq x) = P(X = 0) + P(X = 1) \\ &= 0,2 + 0,16 = 0,36. \end{aligned}$$

When $2 \leq x < 3$ one gets:

$$F(x) = P(X \leq x) = P(X = 0) + P(X = 1) + P(X = 2) = 0,2 + 0,16 + 0,128 = 0,488.$$

When $x \geq 3$ we have $F(x) = 1$.

Thus the cummulative distribution function of X is:

$$F(x) = \begin{cases} 0 & , x < 0 \\ 0,2 & , 0 \leq x < 1 \\ 0,36 & , 1 \leq x < 2 \\ 0,488 & , 2 \leq x < 3 \\ 1 & , 3 \leq x \end{cases}.$$

Remark: Some authors define the cummulative distribution function as $F(x) := P(X < x)$ then the above result looks differently but we think in a similar manner discussing the cases $k < x \leq k+1$.

c) The driver expects to find 2 barriers open because the expected value of X is

$$E(X) = 0 \cdot 0,2 + 1 \cdot 0,16 + 2 \cdot 0,128 + 3 \cdot 0,512 \approx 1.95$$

Problem 4. *The number of buses that arrive at a bus stop in T minutes is a Poisson random variable B with expected value $T/5$.*

a) What is the PMF of B, the number of buses that arrive in T minutes?

b) What is the probability that in a two-minute interval, three buses will arrive?

c) What is the probability of no buses arriving in a 10-minute interval?

d) How much time should you allow so that with probability 0.99 at least one bus arrives?

Solution: a) When something happens at a constant mean rate in a fixed period of time, it usually leads to a mathematical modelling using a Poisson random variable denoted by B with λ, that constant mean rate. In our case we expect $T/5$ buses in a period of length T minutes, thus $\lambda = \frac{T}{5}$ and by its very definition the PMF will be

$$P_B(k) = \begin{cases} \frac{\left(\frac{T}{5}\right)^k}{k!} e^{-\frac{T}{5}}, & \text{if } b \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

b) Choosing $T = 2$ minutes, the probability that three buses arrive in a two minute interval is

$$P_B(3) = \frac{\left(\frac{2}{5}\right)^3}{3!} e^{-\frac{2}{5}} \approx 0.0072$$

c) By choosing T=10 minutes, the probability of zero buses arriving in a ten minute interval is

$$P_B(0) = e^{-\frac{10}{5}} = e^{-2} \approx 0.135 \approx 13\%$$

d) The probability that at least one bus arrives in T minutes is

$$P(B \geq 1) = 1 - P(B = 0) = 1 - e^{-T/5} \geq 0.99$$

Rearranging yields $T \geq 5 \ln 100 \approx 23$ minutes.

Proposed problems

Problem 1. *From a lot of* 100 *items, of which* 10 *are defective a random sample of size* 5 *is selected for quality control. Construct the distribution series of the random number* X *of defective items contained in the sample.*

Problem 2. *A car has four traffic lights on its route. Each of them allows it to move ahead or stop with the probability* 0.5. *Sketch the distribution polygon of the probabilities of the numbers of lights passed by the car before the first stop has occurred.*

Problem 3. *Births in a hospital occur randomly at an average rate of* 1.8 *births per hour. What is the probability of observing* 4 *births in a given hour at the hospital?*

Problem 4. *It is known that* 3% *of the circuit boards from a production line are defective. If a random sample of* 120 *circuit boards is taken from this production line estimate the probability that the sample contains:*

i) Exactly 2 *defective boards.*

ii) At least 2 *defective boards.*

Problem 5. *Four different prizes are randomly put into boxes of cereal. One of the prizes is a free ticket to the local zoo. Suppose that a family of four decides to buy this cereal until obtaining four free tickets to the zoo. What is the probability that the family will have to buy* 10 *boxes of cereal to obtain the four free tickets to the zoo? What is the probability that the family will have to buy* 16 *boxes of cereal to obtain the four free tickets to the zoo?*

Problem 6. *An automatic line in a state of normal adjustment can produce a defective item with probability* p. *The readjustment of the line is made immediately after the first defective item has been produced. Find the average number of items produced between two readjustments of the line.*

Problem 7. *A student takes a multiple-choice test consisting of two problems. The first one has* 3 *possible answers and the second one has* 5. *The student chooses, at random, one answer as the right one from each of the two problems. Find the expected number* $E(X)$ *of right answers* X *of the student. Find the variance* $var(X)$. *Generalize.*

Problem 8. *The number of calls coming per minute into a hotels reservation center is a Poisson random variable of parameter* $\lambda = 3$.

(a) Find the probability that no calls come in a given 1 minute period.

(b) Assume that the number of calls arriving in two different minutes are independent. Find the probability that at least two calls will arrive in a given two minute period.

(c) What is the expected number of calls in a given period of 1 *minute ?*

Problem 9. *As a result of experiments with two devices* A *and* B, *one finds the probability of observing a noise whose level is evaluated in a three-point system:*

Noise level	1	2	3
Device A	0.20	0.06	0.04
Device B	0.06	0.04	0.10

Using the table select the device with lower noise level.

Bibliography

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

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