

# 1 3.05, 10am-12pm

## 1.1 Exercise, normal distribution

MR, 26

Let  $X \sim N(9, 25)$ . Show that  $P(X \leq 0) = 1 - \Phi(9/5)$ , where  $\Phi$  is the cdf of the standard normal variable,  $N(0, 1)$ .

*Hint:* If  $X \sim (\mu, \sigma^2)$ , and  $Z = aX + b$ , then  $Z \sim N(a\mu + b, a^2\sigma^2)$ .

### A solution

Let  $z = x/5 - 9/5$ . Then  $Z \sim N(9/5 - 9/5, 25/25) = N(0, 1)$ .

We calculate next, using  $X = 5Z + 9$ :

$$\begin{aligned} P(X \leq 0) &= P(5Z + 9 \leq 0) \\ &= P(Z \leq -9/5) \\ &= \Phi(-9/5) \text{ by } Z \sim N(0, 1) \text{ and definition of } \Phi \\ &= 1 - \Phi(9/5) \text{ by symmetry of } \Phi \end{aligned}$$

where the symmetry of  $\Phi$  property was used already in lectures.

## 1.2 Exercise, discrete, random vectors

The police department in Berlin is doing computer simulations for how to manage the football hooligans after the upcoming game between FC Union Berlin and Dynamo Dresden. Assume that there will be

A1:  $h \geq 1$  hooligans transferring at Suedkreuz Bahnhof from Olympias-tadion, where the game is, to their train home to Dresden

A2:  $r \geq 1$  police officers to monitor them

A3: the simulation only cares about the relative positions of hooligans and police along a line parallel to the incoming train

A4: the ordering of hooligans and police officers along this line is uniformly random on  $h + r$  elements.

For example, if there were  $h = 5$  hooligans and  $r = 3$  police officers, then the outcomes (elements of  $\Omega$ ), where P represents a police officer, and H a hooligan

- PHHPPHHH, and
- HHHPPHPH

are equally probable.

For  $k = 1, \dots, r - 1$ , define the random variable  $\xi_k$  as

$\xi_k =$  number of hooligans between the  $k^{\text{th}}$  and  $(k + 1)^{\text{st}}$  police officers

where we count from left to right, and define additionally

- $\xi_0$  as the number of hooligans to the left of the first police officer,
  - $\xi_r$  as the number of hooligans to the right of the last police officer.
1. Determine the probability distribution function of the random vector  $\Xi = (\xi_0, \dots, \xi_r)$ , i.e.  $P(\xi_0 = k_0, \dots, \xi_r = k_r)$  as a function of  $k_0, \dots, k_r$ .
  2. For a fixed  $j \in \{0, \dots, r\}$ , determine the marginal probability distribution  $P(\xi_j = k)$ .
  3. Are the component random variables  $\xi_0, \dots, \xi_r$  independent for all  $r$  (and all  $h$ )?

### A solution

**A solution for 1:** Let's consider instead the positions of the police officers from left to right among the  $r + h$  total positions along the line of hooligans and police.

Define  $j_\ell \in \{1, \dots, h + r\}$  as the relative position of the  $\ell^{\text{th}}$  police officer. For example, if we had  $h = 5, r = 3$ , the choice  $j_1 = 1, j_2 = 3, j_3 = 4$  would mean police officers in the 1st, 3rd and 4th slots, i.e. PHHPPHHH.

Such a choice of  $j_1, \dots, j_r$  uniquely determines the  $(\xi_0, \dots, \xi_r)$ , since

$$\begin{aligned}\xi_0 &= j_1 - 1 \\ \xi_1 &= j_2 - j_1 - 1 \\ &\vdots \\ \xi_r &= h + r - j_r\end{aligned}$$

Conversely, to see that a fixed  $\xi_0 = k_0, \dots, \xi_r = k_r$  with  $\sum k_i = h$  uniquely determines a choice of  $j_1, \dots, j_r$ ,  $1 \leq j_1 < \dots < j_r \leq r$ , just solve the above (linear) equations for  $j_1, \dots, j_r$ .

Hence we have a bijection between  $\xi_0 = k_0, \dots, \xi_r = k_r$  with  $\sum k_i = h$  and  $j_1, \dots, j_r$ ,  $1 \leq j_1 < \dots < j_r \leq r$ . Since the order of police officers is unimportant, and each choice of  $j_1, \dots, j_r$ ,  $1 \leq j_1 < \dots < j_r \leq r$  is equally likely (by assumption A3 of uniformly distributed placements of hooligans and police on a line), we have

$$P(\xi_0 = k_0, \dots, \xi_r = k_r) = \frac{1}{\binom{h+r}{r}} = \frac{h!r!}{(h+r)!}$$

if  $k_0 + \dots + k_r = r$ ; otherwise  $P(\xi_0 = k_0, \dots, \xi_r = k_r) = 0$ .

**A solution for 2:** From the lectures we have for discrete probability vectors (aka multivariable discrete random variables)

$$\begin{aligned}P(\xi_j = k) &= \sum_{\substack{k_1, \dots, k_{j-1}, k_{j+1}, \dots, k_r \in \{0, h\}, \\ \sum_{\ell: \ell \neq j} k_\ell = h - k}} P(\xi_1 = k_1, \dots, \xi_j = k, \dots, \xi_r = k_r) \\ &= \frac{h!r!}{(h+r)!} \sum_{\substack{k_1, \dots, k_{j-1}, k_{j+1}, \dots, k_r \in \{0, h\}, \\ \sum_{\ell: \ell \neq j} k_\ell = h - k}} 1\end{aligned}$$

Before continuing, let's look at the concrete case as above with  $h = 5$ ,  $r = 3$  (5 hooligans, 3 police), and consider  $P(\xi_1 = k)$ . The summand on the right now includes terms  $k_2, k_3 \in \{0, 3-1\}$  satisfying  $k_2 + k_3 = h - k$ . In other words, we have one fewer police officer, and  $k$  fewer hooligans, but otherwise the set-up is the same (also the bijection).

In general, the same logic holds for general  $h, r$ , so we get

$$\begin{aligned} P(\xi_j = k_j) &= \frac{h!r!}{(h+r)!} \binom{r+h-k-1}{r-1} \\ &= \frac{h!r!}{(h+r)!} \frac{(r+h-k-1)!}{(r-1)!(h-k)!} \\ &= r \frac{h \cdot (h-1) \dots (h-k+1)}{(r+h)(r+h-1) \dots (r+h-k)} \end{aligned}$$

Note: it is not clear if the final form is prettier than the first line.

**A solution for 3:** To prove that the  $\xi_j$  are *independent*, we must show, for all  $r, h$  (number of police, hooligans, respectively) and for all  $k_1, \dots, k_h$ , that

$$P(\xi_0 = k_0, \dots, \xi_r = k_r) = \prod P(\xi_\ell = k_\ell)$$

To show that they are *not independent* (or dependent), it suffices to find

- a single choice of numbers of hooligans and police  $(h, r)$ ,
- a single pair of indices  $j \neq j'$  and
- a single pair of  $k_j, k_{j'}$  such that

$$P(\xi_j = k_j) \neq P(\xi_{j'} = k_{j'})$$

Let's look first for an example that shows non-independence. In an exam, just go for one. I give multiple counterexamples for pedagogical purposes.

Note that if  $r$  or  $h$  is 0, then all marginal probabilities are also 0, so these cases don't provide a counterexample for independence.

First counterexample (more obvious, less clever): Take  $r = 1, h = 2$  and consider  $(\xi_0 = 0, x_1 = 2)$ . Then our sample (event) space is  $\{HHP, HPH, PHH\}$ , and each outcome is equally probable.

$$P(\xi_0 = 0, x_1 = 2) = 1/3$$

while  $P(\xi_0 = 0) = \frac{2}{3 \cdot 2} = \frac{1}{3}$  and  $P(\xi_1 = 3) = \frac{2}{3 \cdot 2} = \frac{1}{3}$  (you can also see these directly from our above comment about the outcome space and

probabilities of elementary events being all identical), so we have found a case where

$$P(\xi_0 = 0, \xi_1 = 2) = 1/3 \neq P(\xi_0 = 0)P(\xi_1 = 2) = 1/9$$

hence the  $\xi_\ell$  are not in general independent.

Second counterexample (perhaps less obvious, arguably cleverer):

Note that if  $h > 0$ , for  $r \geq 1$ ,  $P(\xi_0 = h, \xi_1 = h) = 0$ , as the  $k_\ell$  values must sum to  $h$ . But

$$\begin{aligned} P(\xi_0 = h) &= \frac{\binom{r-1}{r-1}}{\binom{h+r}{r}} \\ &= P(\xi_1 = h) > 0 \end{aligned}$$

hence the independence equality condition does not hold, and the random variables are dependent.

And the third, arguably cleverest counterexample would be like the first, but taking  $P(\xi_0 = 2, \xi_1 = 2) = 0$ , while each of the marginals  $P(\xi_j) = 1/3$ .

### What about using the multinomial distribution?

Recall, that we define the multinomial distribution as follows. Consider an experiment with  $r$  possible outcomes (a physical model was randomly throwing balls into  $r$  buckets), and let  $p_i$  be the probability of outcome  $i$ , the  $p_i$  must satisfying

$$\begin{aligned} 0 \leq p_i \leq 1 \text{ for each } i \\ \sum_{i=1}^r p_i = 1 \end{aligned}$$

Assume further that we perform  $n$  trials of this experiment independently, and define the random variable  $X_i$  as the counts of outcome  $i$  occurring.

The *multinomial distribution* is the distribution of the random vector

$$X = (X_1, \dots, X_r)$$

You proved in the lecture that, under the above assumptions about independence,

$$P(X_1 = k_1, \dots, X_r = k_r) = \frac{n!}{k_1! \cdot \dots \cdot k_r!} p_1^{k_1} \dots p_r^{k_r}$$

if  $k_i \in \mathbb{N}$ ,  $\sum k_i = n$ , and 0 otherwise.

*Geometry side note:* The solutions to the above equations on the  $p_i$  define a polytope in  $R^r$ , as they represent the solution space of  $2 \times r$  linear inequalities (the conditions on the  $p_i$ ) and one linear equality (total probability sums to 1).

A few of you noticed that our count-the-hooligans-between-police exercise seems a natural fit for the multinomial distribution. First, let's adapt the notation to fit our exercise setup. It's a bit subtle, but we'll see that the multinomial distribution does not precisely fit our situation.

In the police-hooligan situation, we would model the number of hooligans  $\xi_0, \dots, \xi_r$  as the counts of hooligans among  $r + 1$  buckets, with the  $0^{th}$  bucket being the hooligans to the left of the first police officer, the  $r^{th}$  bucket being the hooligans to the right of the final,  $r^{th}$  officer, and the remaining  $\xi_j$  being the count of hooligans between the  $j^{th}$  and  $(j + 1)^{st}$  officer.

If we knew the  $p_i$ ,  $0 \leq i \leq r$ , then the solution to our problem would be

$$P(\xi_0 = k_0, \dots, \xi_r = k_r) = \frac{h!}{k_0! \cdot \dots \cdot k_r!} p_0^{k_0} \dots p_r^{k_r}$$

It's tempting to now take the "uniform" assumption of the problem description to set  $p_i = 1/(r + 1)$ , saying that landing in each bin (of police officers) is the same, which would give us a probability distribution depending non-trivially on the  $k_0, \dots, k_r$ , in contradiction of our solution above.

The first critical piece is that assuming fixed values for  $p_i$  (whether equal to  $1/(r + 1)$  for all, or something different for each) implies that the relative positions of the police officers on the line (which has  $h + r$

places) are already fixed.

The second critical difference is that the multinomial setup assumes a sequence of distinguishable trials or experiments (in case of throwing balls into buckets, a sequence of trials), whereas our police hooligan situation (as well as the original formulation of a well-shuffled deck of  $h$  blue cards and  $r$  red cards) assume indistinguishable trials.

In other terms, the multinomial distribution applies to a sequence of distinguishable trials with  $r$  possible outcomes for each trial, each outcome having a fixed probability, while our exercise set-up applies to a single trial (or sample) from a space of configurations.

*Physics side note:* In statistical physics, these two types of distributions go under the heading of *Maxwell-Boltzmann*, or *classical*, statistics for distinguishable, interactive particles (multinomial) vs the \*Bose-Einstein\*, or \*quantum\*, statistics for indistinguishable, non-interacting particles.

This sentence from the Wikipedia page about quantum statistics highlights the key difference for our question: In quantum statistics, “interchanging any two particles does not lead to a new configuration of the system”. Swapping hooligans or police officers doesn’t change our configuration.

*Physics and geometry aside:* Maxwell-Boltzmann statistics assume that the particles satisfy the Boltzmann exclusion principle of no two particles being allowed to occupy the same physical space at the same time, whereas the Bose-Einstein statistics allow overlapping particles (like a Bose-Einstein condensate). Informally, our (Bose-Einstein) space of configurations of police officers and hooligans allow an arbitrary number of hooligans (limited only by the total number of hooligans) to fit into the space defined by officer positions, while a multinomial, Maxwell-Boltzmann setup would impose a fixed number of spots hooligans could fit into determined by the fixed positions of the police officers.

*Hollywood aside:* As far as I know, being familiar with Marvel’s Quantum Realm / Multiverse movies, e.g. *Ant Man*, yield no practical benefit to understanding actual quantum statistics.

### 1.3 Exercise, hat (or jacket check) problem

$n$  visitors to the Komische Opera ('funny opera') in Berlin all check in black, Jack Wolfskin jackets before a performance of Mozart's Don Giovanni. When they come to pick up their jackets after the performance, with uniform probability, the coat check clerk gives them a random black, Jack Wolfskin jacket.

Let  $X$  be the random variable defined as the number of visitors who get the correct jacket back, so  $X \in \{0, \dots, n\}$ .

On average, how many visitors do we expect to get back their own jacket?

*Hint:* Use indicator functions / variables, and expected value.

*Cultural side notes:*

1. I may have seen Ralf Fiennes, of Harry Potter Voldemort and English Gardner fame, waiting for a performance of Don Giovanni to start in Berlin back in 2008.
2. A large proportion of Germans wear black, Jack Wolfskin jackets. After a work party once when I went to pick up my jacket, I jokingly tried to help the coat clerk by telling her that mine was the black, Jack Wolfskin jacket.

#### A solution

Let  $J_k$  be the event that the  $k^{\text{th}}$  visitor gets their jacket back. By assumption of uniformity,  $P(J_k) = 1/n$ .

Define the indicator function  $I_k$  as

$$I_k = 1 \text{ if } J_k \text{ occurs; else } 0$$

In other words,  $I_k$  is the indicator function for the  $k^{\text{th}}$  person getting their correct jacket back.

Since  $X = I_1 + \dots + I_n$ , we can use linearity of expectation to calculate

$$E(X) = E\left(\sum_{j=1}^n I_j\right) = \sum_{j=1}^n E(I_j)$$



but  $E(I_j) = 1 \cdot 1/n + 0 \cdot (n-1)/n = 1/n$  for each  $j$ , so

$$E(X) = n \cdot 1/n = 1$$

Hence we expect one person to get their correct jacket back.

*Mathematical aside:* Returning to the question about the police-officer and hooligan configurations, in this exercise, the trials in question (picking up your jacket) are *distinguishable*, as the setup requires that the similar looking jackets be identifiable as belonging to one owner.

## 1.4 Exercise: A variant on the blue and red pills

(Adapted from Stirzaker, Probability and Random Variables, Chapter 5, ex. 11)

In the movie the Matrix, the character Morpheus offers another character Neo two pills. If Neo chooses the blue pill, he will return to his old life. If he chooses the red one, his eyes will be opened to some truth, and he will join Morpheus' band.

In this variant (also of Pólyeva's urn, see MR, vaja 4), Neo has to choose blindly from a bucket initially containing one red and one blue pill. If he picks red, he must swallow it and the game is over. If he picks blue, then he returns the blue pill and Morpheus adds an extra red pill.

Let  $X$  be the number of draws. Calculate

1. The distribution  $P(X > k)$
2.  $E(X)$  (hint: use the method of indicators)

### A solution for 1, $P(X = k)$

This is a geometric like process (a number of successive failures followed by a single success), except that the probability of success or failure changes each time.

Now suppose  $k \in \{1, \dots, 4\}$ . By Morpheus' rules,  $X > k$  means that the first  $k$  draws were blue

$$P(X > k) = \prod_{j=1}^k (j+1) = \frac{1}{(k+1)!}$$

**A solution with indicator functions for 2,  $E(X)$**

Let  $I_j$  be the event that Neo draws again at step  $j$ . By construction,  $X = \sum_{j=1}^{\infty} I_j$ , and by linearity of expectation,

$$\begin{aligned} E(X) &= \sum_{j=1}^{\infty} E(I_j) \\ &= \sum_{j=1}^{\infty} P(X \geq j) \\ &= \sum_{j=1}^{\infty} P(X > j-1) = \sum_{j=1}^{\infty} \frac{1}{j!} \end{aligned}$$

But  $\sum_{j=1}^{\infty} \frac{1}{j!} = e - 1 \sim 1.718$ , so  $E(X) = e - 1$ .