Probability distributions
Lectures, questions only

Nicky van Foreest and Ruben van Beesten

March 1, 2021

CONTENTS

1 LECTURE 1

Ex 1.1. Consider 12 football players on a football field. Eleven of them are players of FC Barcelone, the other one is an arbiter. We select a random player, uniform. This player must take a penalty. The probability that a player of Barcelone scores is 70%, for the arbiter it is 50%. Let $P \in \{A,B\}$ be r.v that corresponds to the selected player, and $S \in \{0,1\}$ be the score.

1. What is the PMF? In other words, determine $\mathsf{P}\{P = B, S = 1\}$ and so on for all possibilities.
2. What is $\mathsf{P}\{S = 1\}$? What is $\mathsf{P}\{P = B\}$?
3. Show that S and P are dependent.

An insurance company receives on a certain day two claims $X, Y \geq 0$. We will find the PMF of the loss $Z = X + Y$ under different assumptions.

The joint CDF $F_{X,Y}$ and joint PMF $p_{X,Y}$ are assumed known.

Ex 1.2. Why is it not interesting to consider the case $\{X = 0, Y = 0\}$?

Ex 1.3. Find an expression for the PMF of $Z = X + Y$.

Suppose $p_{X,Y}(i,j) = c\, I_{i=j} I_{1\leq i\leq 4}$.

Ex 1.4. What is c?

Ex 1.5. What is $F_X(i)$? What is $F_Y(j)$?

Ex 1.6. Are X and Y dependent? If so, why, because $1 = F_{X,Y}(4,4) = F_X(4)F_Y(4)$?

Ex 1.7. What is $\mathsf{P}\{Z = k\}$?

Ex 1.8. What is $\mathsf{V}[Z]$?

Now take X, Y iid $\sim$ Unif($\{1,2,3,4\}$) (so now no longer $p_{X,Y}(i,j) = I_{i=j} I_{1\leq i\leq 4}$).

Ex 1.9. What is $\mathsf{P}\{Z = 4\}$?

Remark 1.1. We can make lots of variations on this theme.

1. Let $X \in \{1,2,3\}$ and $Y \in \{1,2,3,4\}$.
2. Take $X \sim$ Pois(λ) and $Y \sim$ Pois(μ). (Use the chicken-egg story)
3. We can make X and Y such that they are (both) continuous, i.e., have densities. The conceptual ideas[1] don't change much, except that the summations become integrals.
4. Why do people often/sometimes (?) model the claim sizes as iid $\sim$ Norm(μ, σ^2)? There is a slight problem with this model (can real claim sizes be negative?), but what is the way out?

[1] Unless you start digging deeper. Then things change drastically, but we skip this technical stuff.

5. The example is more versatile than you might think. Here is another interpretation.

   A supermarket has 5 packets of rice on the shelf. Two customers buy rice, with amounts X and Y. What is the probability of a lost sale, i.e., $\mathsf{P}\{X+Y>5\}$? What is the expected amount lost, i.e., $\mathsf{E}[\max\{X+Y-5,0\}]$?

   Here is yet another. Two patients arrive in to the first aid of a hospital. They need X and Y amounts of service, and there is one doctor. When both patients arrive at 2 pm, what is the probability that the doctor has work in overtime (after 5 pm), i.e., $\mathsf{P}\{X+Y>5-2\}$?

HERE IS SOME extra material for you practice on 2D integration.

Ex 1.10. We have a continuous r.v. $X \geq 0$ with finite expectation. Use 2D integration and indicators to prove that

$$\mathsf{E}[X] = \int_0^\infty x f(x)\,\mathrm{d}x = \int_0 G(x)\,\mathrm{d}x, \tag{1.1}$$

where $G(x)$ is the survival function.

Ex 1.11. Explain that for a continuous r.v. X with CDF F and a and b (so it might be that $a>b$),

$$\mathsf{P}\{a<X<b\} = [F(b)-F(a)]^+. \tag{1.2}$$

INDICATORS ARE GREAT functions, and I suspect you underestimated the importance of these functions. They help to keep your formulas clean. You can use them in computer code as logical conditions, or to help counting relevant events, something you need when numerically estimating multi-D integrals for machine learning for instance. And, even though I(=NvF) often prefer to use figures over algebra to understand something, when it comes to integration (and reversing the sequence of integration in multiple integrals) I find indicators easier to use.

Moreover, you should know that in fact, *expectation* is the fundamental concept in probability theory, and the *probability of an event is defined* as

$$\mathsf{P}\{A\} := \mathsf{E}[I_A]. \tag{1.3}$$

Thus, the fundamental bridge is actually an application of LOTUS to indicator functions. REREAD BH.4.4!

Ex 1.12. What is $\int_{-\infty}^{\infty} I_{0\leq x\leq 3}\,\mathrm{d}x$?

Ex 1.13. What is

$$\int x I_{0\leq x\leq 4}\,\mathrm{d}x? \tag{1.4}$$

When we do an integral over a 2D surface we can first integrate over the x and then over the y, or the other way around, whatever is the most convenient. (There are conditions about how to handle multi-D integral, but for this course these are irrelevant.)

Ex 1.14. What is

$$\iint xy I_{0\leq x\leq 3} I_{0\leq y\leq 4}\,\mathrm{d}x\,\mathrm{d}y? \tag{1.5}$$

Ex 1.15. What is

$$\iint I_{0\leq x\leq 3} I_{0\leq y\leq 4} I_{x\leq y}\,\mathrm{d}x\,\mathrm{d}y? \tag{1.6}$$

Ex 1.16. Take $X \sim \text{Unif}([1,3]), Y \sim \text{Unif}([2,4])$ and independent. Compute

$$\mathsf{P}\{Y \leq 2X\}. \tag{1.7}$$

2 LECTURE 2

Read the problems of `memoryless\_excursions.pdf`. All the problems in that document relate to topics discussed in Sections BH.7.1 and BH.7.2, and quite a lot of topics you have seen in the previous course on probability theory.

3 LECTURE 3

Ex 3.1. We ask a married woman on the street her height X. What does this tell us about the height Y of her spouse? We suspect that taller/smaller people choose taller/smaller partners, so, given X, a simple estimator $\hat{Y}$ of Y is given by

$$\hat{Y} = aX + b.$$

(What is the sign of a if taller people tend to choose taller people as spouse?) But how to determine a and b? A common method is to find a and b such that the function

$$f(a,b) = \mathsf{E}\left[(Y - \hat{Y})^2\right]$$

is minimized. Show that the optimal values are such that

$$\hat{Y} = \mathsf{E}[Y] + \rho\frac{\sigma_Y}{\sigma_X}(X - \mathsf{E}[X]),$$

where ρ is the correlation between X and Y and where σ_X and σ_Y are the standard deviations of X and Y respectively.

Ex 3.2. Using scaling laws often can help to find errors. For instance, the prediction $\hat{Y}$ should not change whether we measure the height in meters or centimetres. In view of this, explain that

$$\hat{Y} = \mathsf{E}[Y] + \rho\frac{\mathsf{V}[Y]}{\sigma_X}(X - \mathsf{E}[X])$$

must be wrong.

Ex 3.3. N people throw their hat in a box. After shuffling, each of them takes out a hat at random. How many people do you expect to take out their own hat (i.e., the hat they put in the box); what is the variance? In BH.7.46 you have to solve this analytically. In the exercise here you have to write a simulator for compute the expectation and variance.

4 LECTURE 4

Ex 4.1. BH.7.65 Let $(X_1,\ldots,X_k)$ be Multinomial with parameters n and $(p_1,\ldots,p_k)$. Use indicator r.v.s to show that $\mathrm{Cov}(X_i,X_j) = -np_ip_j$ for $i \neq j$.

Ex 4.2. Suppose (X,Y) are bivariate normal distributed with mean vector $\mu = (\mu_X,\mu_Y) = (0,0)$, standard deviations $\sigma_X = \sigma_Y = 1$ and correlation ρ_{XY} between X and Y. Specify the joint pdf of X and $X+Y$.

The following exercises will show how probability theory can be used in finance. We will look at the tradeoff between risk and return in a financial portfolio.

John is an investor who has $10,000 to invest. There are three stocks he can choose from. The returns on investment (A,B,C) of these three stocks over the following year (in terms of percentages) follow a multinomial distribution. The expected returns on investment are $\mu_A = 7.5\%$, $\mu_B = 10\%$, $\mu_C = 20\%$. The corresponding standard deviations are $\sigma_A = 7\%$, $\sigma_B = 12\%$ and $\sigma_C = 17\%$. Note that risk (measured in standard deviation) increases with expected return. The correlation coefficients between the different returns are $\rho_{AB} = 0.7$, $\rho_{AC} = -0.8$, $\rho_{BC} = -0.3$.

Ex 4.3. Suppose the investor decides to invest $2,000 in stock A, $4,000 in stock B, $2,000 in stock C and to put the remaining $2,000 in a savings account with a zero interest rate. What the expected value of his portfolio after a year?

Ex 4.4. What is the standard deviation of the value of the portfolio in a year?

Ex 4.5. John does not like losing money. What is his probability of having made a net loss after a year?

John has a friend named Mary, who is a first-year EOR student. She has never invested money herself, but she is paying close attention during the course Probability Distributions. She tells her friend: “John, your investment plan does not make a lot of sense. You can easily get a higher expected return at a lower level of risk!”

Ex 4.6. Show that Mary is right. That is, make a portfolio with a higher expected return, but with a lower standard deviation.
Hint: Make use of the ***negative correlation*** *between C and the other two stocks!*

5 LECTURE 5

HERE IS A NICE geometrical explanation of how the normal distribution originates.

Ex 5.1. Suppose $z_0 = (x_0, y_0)$ is the target on a dart board at which Barney (our national darts hero) aims, but you can also interpret it as the true position of a star in the sky. Let z be the actual position at which the dart of Barney lands on the board, or the measured position of the star. For ease, take z_0 as the origin, i.e., $z_0 = (0,0)$. Then make the following assumptions:

1. The disturbance (x, y) has the same distribution in any direction.
2. The disturbance (x, y) along the x direction and the y direction are independent.
3. Large disturbances are less likely than small disturbances.

Show that the disturbance along the x-axis (hence y-axis) is normally distributed. You can use BH.8.17 as a source of inspiration. (This is perhaps a hard exercise, but the solution is easy to understand and very useful to memorize.)

We next find the normalizing constant of the normal distribution (and thereby offer an opportunity to practice with change of variables).

Ex 5.2. For this purpose consider two circles in the plane: $C(N)$ with radius N and $C(\sqrt{2}N)$ with radius $\sqrt{2}N$. It is obvious that the square $S(N) = [-N, N] \times [-N, N]$ contains the first circle, and is contained in the second. Therefore,

$$\iint_{C(N)} f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y \leq \iint_{S(N)} f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y \leq \iint_{C(\sqrt{2}N)} f_{X,Y}(x, y)\,\mathrm{d}x\,\mathrm{d}y. \tag{5.1}$$

Now substitute the normal distribution of [5.1]. Then use polar coordinates (See BH.8.1.9) to solve the integrals over the circles, and derive the normalization constant.

BENFORD'S LAW MAKES a statement on the first significant digit of numbers. Look it up on the web; it is a fascinating law. It's used to detect fraud by insurance companies and the tax department, but also to see whether the US elections in 2020 have been rigged, or whether authorities manipulate the statistics of the number of deceased by Covid. You can find the rest of the analysis in Section 5.5 of 'The art of probability for scientists and engineers' by R.W. Hamming. The next exercise is a first step in the analysis of Benford's law.

Ex 5.3. Let X, Y be iid with density f and support $[1, 10)$. Find an expression for the density of $Z = XY$. What is the support (domain) of Z? If $X, Y \sim \mathrm{Unif}([1, 10))$, what is f_Z?

6 LECTURE 6

Ex 6.1. Let $X \sim \text{Pois}(\lambda)$ and $Y \sim \text{Pois}(\mu)$ be independent. What is the distribution of $Z = X + Y$?

Ex 6.2. (BH.8.4.3.) Let $X_1, X_2, \ldots$ be i.i.d. $\text{Exp}(\lambda)$ distributed. Let $T_n = \sum_{k=1}^{n} X_k$. Show that T_n has the following pdf:

$$f_{T_n}(t) = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}, \quad t > 0. \tag{6.1}$$

That is, show that T_n follows a *Gamma distribution* with parameters n and λ. (We will learn about the Gamma distribution in BH.8.4.)

Ex 6.3. Let X, Y be i.i.d. $\mathcal{N}(0,1)$ distributed and define $Z = X + Y$. Show that $Z \sim \mathcal{N}(0,2)$ using a convolution integral.

7 LECTURE 7

Ex 7.1. At the end of Story 2 of Bayes' billiards (BH.8.3.2) there is the expression

$$\beta(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}. \tag{7.1}$$

Derive this equation.

Ex 7.2. In the Beta-Binomial conjugacy story, BH take as prior $f(p) = I_{p\in[0,1]}$, and then they remark that when $f(p) \sim \beta(a,b)$ for general $a,b \in \mathbb{N}$, we must obtain the negative hypergeometric distribution. I found this pretty intriguing, so my question is: Relate the Bayes' billiards story to the story of the Negative Hypergeometric distribution, and, in passing, provide an interpretation of a and b in terms of white and black balls. Before trying to answer this question, look up the details of the negative hypergeometric distribution. (In other words, this exercise is meant to help you sort out the details of the remark of BH about the negative hypergeometric distribution.)

The next real exercise is about recursion applied to the negative hypergeometric distribution. But to get in the mood, here is short fun question on how to use recursion.

Ex 7.3. We have a chocolate bar consisting of n small squares. The bar is in any shape you like, square, rectangular, whatever. What is the number of times you have to break the bar such that you end up with the n single pieces?

Ex 7.4. Use recursion to find the expected number X of black balls drawn without replacement at random from an urn containing $w \geq 1$ white balls and b black balls before we draw 1 white ball. In other words, I ask to use recursion to compute $\mathsf{E}[X]$ for X a negative hypergeometric distribution with parameters $w, b, r = 1$ and show that

$$\mathsf{E}[X] = \frac{b}{w+1} \tag{7.2}$$

Ex 7.5. Extend the previous exercise to cope with the case $r \geq 2$. For this, write $N_r(w,b)$ for an urn with w white balls and b black balls, and r white balls to go.

8 LECTURE 8

Ex 8.1. Let X be a continuous random variable with a pdf

$$f_X(x) = \begin{cases} c, & \text{if } 0 \le x \le 4, \\ 0, & \text{otherwise.} \end{cases} \tag{8.1}$$

1. What is the value of c?
2. What is the distribution of X?
3. Do we need to know the value of c to determine the distribution of X?

Ex 8.2. Let X be a continuous random variable with a pdf

$$f_X(x) = c \cdot e^{-\frac{(x-4)^2}{8}}, \quad x \in \mathbb{R}. \tag{8.2}$$

1. What is the value of c?
2. What is the distribution of X?
3. Do we need to know the value of c to determine the distribution of X?

(BH.8.4.5) Fred is waiting at a bus stop. He knows that buses arrive according to a Poisson process with rate λ buses per hour. Fred does not know the value of λ, though. Fred quantifies his uncertainty of λ by the *prior distribution* $\lambda \sim \text{Gamma}(r_0, b_0)$, where r_0 and b_0 are known, positive constants with r_0 an integer.

Fred has waited for t hours at the bus stop. Let Y denote the number of buses that arrive during this time interval. Suppose that Fred has observed that $Y = y$.

Ex 8.3. Find Fred's (hybrid) joint distribution for Y and λ.

Ex 8.4. Find Fred's marginal distribution for Y. Use this to compute $\mathsf{E}[Y]$. Interpret the result.

Ex 8.5. Find Fred's posterior distribution for λ, i.e., his conditional distribution of λ given the data y.

Ex 8.6. Find Fred's posterior mean $\mathsf{E}[\lambda|Y = y]$ and variance $\mathsf{V}[\lambda|Y = y]$.

9 LECTURE 9

Ex 9.1. The lifetime X of a machine is $\text{Exp}(\lambda)$. Compute $\mathsf{E}[X \mid X \leq \tau]$ where τ is some positive constant.

Define $Y = I_{X \leq \tau}$. Use Adam's law and LOTP to show that $\mathsf{E}[X] = \mathsf{E}[\mathsf{E}[X \mid Y]] = 1/\lambda$. Observe that this is just a check on our results.

Ex 9.2. We have a station with two machines, one is working, the other is off. If the first fails, the other machine takes over. The repair time of the first machine is a constant τ. If the second machine fails before the first is repaired, the station stops working, i.e., is 'down'. Use a conditioning argument to find the expected time $\mathsf{E}[T]$ until the station is down when the lifetimes of both machines is iid $\sim \text{Exp}(\lambda)$.

Ex 9.3. We draw, with replacement, balls, numbered 1 to N, from an urn. Find a recursion to compute the expected number $\mathsf{E}[T]$ of draws necessary to see all balls.

Ex 9.4. Write code to compute $\mathsf{E}[T]$ for $N = 45$.

Ex 9.5. We draw, with replacement, balls, numbered 1 to N, from an urn, but 6 at a time (not just one as in the previous exercise). Find a recursion to compute the expected number $\mathsf{E}[T]$ of draws necessary to see all balls.

Ex 9.6. For the previous exercise, compute $\mathsf{E}[T]$ for $N = 45$.

10 LECTURE 10

We have a wooden stick of length 100 cm that we break twice. First, we break the stick at a random point that is uniformly distributed over the entire stick. We keep the left end of the stick. Then, we break the remaining stick again at a random point, uniformly distributed again, and we keep the left end again.

Ex 10.1. What is the expected length of the stick we end up with?

Ex 10.2. Now we change the story slightly. Every time we break a stick, we keep the *longest* part. What is the expected length of the remaining stick?

(Same story as last week; BH.8.4.5) Fred is waiting at a bus stop. He knows that buses arrive according to a Poisson process with rate λ buses per hour. Fred does not know the value of λ, though. Fred quantifies his uncertainty of λ by the *prior distribution* $\lambda \sim \text{Gamma}(r_0, b_0)$, where r_0 and b_0 are known, positive constants with r_0 an integer.

Fred has waited for t hours at the bus stop. Let Y denote the number of buses that arrive during this time interval. Suppose that Fred has observed that $Y = y$.

Ex 10.3. How many buses does Fred expect to observe? I.e., compute $\mathsf{E}[Y]$

(BH.9.3.10). An extremely widely used method for data analysis in statistics is *linear regression*. In its most basic form, the linear regression model uses a single explanatory variable X to predict a response variable Y. For instance, let X be the number of hours studied for an exam and let Y be the grade on the exam. The linear regression model assumes that the conditional expectation of Y is *linear* in X:

$$\mathsf{E}[Y|X] = a + bX. \tag{10.1}$$

Ex 10.4. Show that an equivalent way to express this is to write

$$Y = a + bX + \varepsilon, \tag{10.2}$$

where ε is a random variable (called the *error*) with $\mathsf{E}[\varepsilon|X] = 0$.

Ex 10.5. Solve for the constants a and b in terms of $\mathsf{E}[X]$, $\mathsf{E}[Y]$, $\mathsf{Cov}[X,Y]$, and $\mathsf{V}[Y]$.

11 HINTS

h.1.10. Check the proof of BH.4.4.8

h.1.11. Recall that $F \in [0,1]$.

h.6.1. Use the Binomial theorem:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, \tag{11.1}$$

for any nonnegative integer n.

h.6.2. Use mathematical induction.

h.7.4. For ease, write $N(w,b) = \mathsf{E}[X]$ for an urn with $w \geq r = 1$ white balls and b black balls. Then explain that

$$N(w,0) = 0 \quad \text{for all } w, \tag{11.2}$$

$$N(w,b) = \frac{b}{w+b}(1+N(w,b-1)). \tag{11.3}$$

Then show that this implies that $N(w,b) = b/(w+1)$.

h.7.5. Explain that $N_0(w,b) = 0$ and

$$N_r(w,b) = \frac{w}{w+b}N_{r-1}(w-1,b) + \frac{b}{w+b}(1+N_r(w,b-1)). \tag{11.4}$$

Then show that this implies that $N_r(w,b) = rb/(w+1)$.