NMAI059 – Probability and Statistics 1

Mykhaylo Tyomkyn

Lecture 3 - Discrete random variables.

Often we are interested not in the actual outcome of a random experiment, but in some function of it.

Example 1 *Consider the gambler's ruin problem from Lecture 2. We may want to track the time (number of tosses) it takes for the game to end.*

Example 2 *We throw a dart at a circular target. We are interested not in the exact place it hits, but in its distance from the centre.*

Definition 1 (Discrete random variable) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X: \Omega \to \mathbb{R}$ is a* (real-valued) discrete random variable *if the set $\operatorname{Im} X = \{X(\omega): \omega \in \Omega\}$ is countable* [1] *[and for every $x \in \mathbb{R}$ we have $\{\omega \in \Omega: X(w) = x\} \in \mathcal{F}$].*

In particular, when Ω is countable, every function $X: \Omega \to \mathbb{R}$ is a discrete random variable. [2]

Definition 2 (Probability mass function) *The* probability mass function (pmf) *of a discrete random variable $X: \Omega \to \mathbb{R}$ is the function $p_X: \mathbb{R} \to [0,1]$, $\quad p_X(x) := \mathbb{P}(X = x)$.* [3]

Note that, by σ-additivity, we always have

$$\sum_{x \in \operatorname{Im} X} p_X(x) = \sum_{x \in \operatorname{Im} X} \mathbb{P}(X = x) = 1,$$

as the sets $\{\omega \in \Omega: X(\omega) = x\}$ form a countable partition of Ω. Therefore, with $S = \operatorname{Im} X$ and $Q: 2^S \to [0,1]$, $Q(A) = \sum_{x \in A} p_X(x)$ we obtain a countable probability space $(S, 2^S, Q)$. In other words, random variables can be used as a tool for constructing probability spaces. Consequently we shall speak of *distributions* of discrete random variables, referring to their pmf's.

Let us now introduce some of the best-known discrete random variables and their distributions.

Example 3 (Bernoulli distribution) *Consider a coin that comes up 'heads' with probability $p \in [0,1]$. Let $X: \Omega = \{H, T\} \to \mathbb{R}$ count the number of 'heads' in a single toss. That is, $X(H) = 1; X(T) = 0$. Then we have $p_X(1) = p$ and $p_X(0) = 1 - p$. This probability function on $\{0,1\}$ is known as the* (p-)Bernoulli distribution*, denoted $Bern(p)$, and X is a (p-)Bernoulli random variable. We write $X \sim Bern(p)$ to signify that $p_X = Bern(p)$.* [4]

[1] The last condition can be ignored if we assume $\mathcal{F} = 2^\Omega$

[2] By convention we shall use letters A, B, C for events, X, Y, Z for random variables and x, y, z for their values

[3] Another convention is to denote events involving random variables by their properties. For instance, $\mathbb{P}(X = x)$ is a convenient shorthand for $\mathbb{P}(\{\omega \in \Omega: X(w) = x\})$.

[4] And we shall use $\sim$ similarly for other 'named' distributions.

Example 4 (Binomial distribution) *Let now X be the number of 'heads' in n independent tosses of the coin from the previous example. Then, with $\Omega = \{H, T\}^n$, for each $k \in \{0, 1 \dots, n\}$ we have*

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

This is known as the binomial distribution *with parameters n and p, in notation $X \sim Bin(n, p)$. Note that by the binomial theorem $\sum_{k=0}^{n} p_X(k) = (p + (1-p))^n = 1$, as it should.*

Just like functions in general, random variables on the same space can be added and multiplied: if $X, Y : \Omega \to \mathbb{R}$ are discrete random variables, then so are $X + Y$ and XY (where $(X + Y)(\omega) = X(\omega) + Y(\omega)$, and analogously for the product[5]). Similarly one can define arbitrary binary operations on random variables, such as $\max(X, Y), \min(X, Y), X^Y$ etc.

In the same fashion, if X is a discrete random variable and $f : \mathbb{R} \to \mathbb{R}$ a function, then $f(X)$ is a discrete random variable. This allows us to speak of $X^2, \sqrt{X}, e^X$ etc.

Intuitively, a $Bin(n, p)$ random variable is the sum of n p-Bernoulli random variables, but one needs to be careful describing the underlying probability space. We will return to this question when discussing the joint distribution and independence of random variables.

Alongside the pmf, there is another important function related to X.

Definition 3 (Cumulative distribution function) *Let $X : \Omega \to \mathbb{R}$ be a discrete random variable. Then*

$$F_X : \mathbb{R} \to [0, 1], \quad F_X(x) := \mathbb{P}(X \leq x)$$

is called the cumulative distribution function (cdf) *of X.*

Exercise 1 *Determine F_X, for $X \sim Bern(p)$.*

Note that we can express the cdf using the pmf

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{t \leq x} \mathbb{P}(X = t) = \sum_{t \leq x} p_X(t).$$

In the opposite direction, expressing the pmf using the cdf is in general not straightforward (can you tell why?) but when $\operatorname{Im} X \subseteq \mathbb{Z}$ we have $p_X(x) = F_X(x) - F_X(x - 1)$.

Theorem 1 (Properties of the cdf) *For any discrete random variable X we have*

1. *F_X is non-decreasing.*
2. *$\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.*
3. *F_X is right-continuous: $\lim_{\epsilon \to 0^+} f(x + \epsilon) = f(x)$ for all $x \in \mathbb{R}$.*

Proof Assertion 1. follows by monotonicity and assertions 2. and 3. by continuity of probability (exercise). □

Let us now go back to considering examples of discrete random variables and their distributions.

[5] Exercise: Check that $X + Y$ takes countably many values if X and Y do.

Example 5 (Geometric distribution) *Let $0 < p < 1$. Take a coin that comes up 'heads' with probability p and toss it repeatedly. Let X be the 'time' (toss number) when the coin comes up 'heads' for the first time. Then, for each $n \in \mathbb{N}$ we have*

$$\mathbb{P}(X = n) = p(1-p)^{n-1}.$$

We say that X is geometrically distributed *with parameter p. In notation, $X \sim Geo(p)$. Note that*

$$\sum_{n \geq 1} \mathbb{P}(X = n) = p(1 + (1-p) + (1-p)^2 + \dots) = p \cdot \frac{1}{1-(1-p)} = 1,$$

indeed. Note also that $\mathbb{P}(X > n) = (1-p)^n$, and so

$$F_X(n) = 1 - (1-p)^n.$$

Exercise 2 *Compare this with the geometric distribution example from Lecture 1, and verify that the latter was $Geo(1/2)$.*

Example 6 (Hypergeometric distribution) *Suppose an urn contains N balls, of which K are white and $N-K$ are black. We draw n balls uniformly at random, without replacement. Let X count the number of white balls drawn, then*

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}. \tag{1}$$

This is known as the hypergeometric distribution *and denoted $X \sim Hyper(N, K, n)$.*

Remark 1 *We can use this fact to prove the combinatorial identity*

$$\sum_{k=0}^{n} \binom{K}{k}\binom{N-K}{n-k} = \binom{N}{n}.$$

Note that if instead we draw the balls with replacement, the distribution of random variable Y measuring the same quantity (number of black balls drawn) will be $Bin(n, K/N)$, i.e.,

$$p_Y(k) = \binom{n}{k}\left(\frac{K}{N}\right)^k \left(1 - \frac{K}{N}\right)^{n-k}.$$

This is a much more easy-to-handle expression than (1), and not too different in value in most cases. We therefore will try to replace the hypergeometric model with the binomial one whenever we can.

Example 7 (Poisson distribution) *Let $\lambda > 0$ be a real number. Define the* Poisson distribution *$Pois(\lambda)$ to be the probability function $\mathbb{P}$ on $\{0, 1, 2, \dots,\}$ defined via*

$$\mathbb{P}(\{k\}) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}.$$

Note that this indeed defines a probability function, as

$$\sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

Unlike in the previous examples there is no completely obvious way of exhibiting a Poisson random variable in a simple coin or urn experiment. However, Poisson random variables are ubiquitous in Nature, as we shall see shortly.

Theorem 2 (Poisson approximation of the binomial distribution) *Let $\lambda > 0$ be fixed and for every $n \in \mathbb{N} \cap (\lambda, \infty)$ let $X_n \sim Bin(n, \lambda/n)$. Then for any fixed $k \in \{0, 1, 2, \dots\}$ we have*

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Proof Using that (for k fixed and $n \to \infty$) we have $\binom{n}{k} = (1 + o(1))\frac{n^k}{k!}$, $(1 - \lambda/n)^k \to 1$ and $(1 - \lambda/n)^n \to e^{-\lambda}$, we obtain

$$\begin{aligned} \mathbb{P}(X_n = k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1 - \frac{\lambda}{n}\right)^{n-k} = (1 + o(1))\frac{n^k}{k!}\frac{\lambda^k}{n^k}\left(1 - \frac{\lambda}{n}\right)^n \\ &= (1 + o(1))\frac{\lambda^k}{k!}\left(1 - \frac{\lambda}{n}\right)^n = (1 + o(1))\frac{\lambda^k}{k!}e^{-\lambda} \end{aligned}$$

□

So, suppose now that we want to estimate X, the number of meteorites (say, of minimum size 1 cubic meter) hitting our planet within a year (time interval T). To simulate this, it makes sense to discretize the model by dividing T into a large number of n equal subintervals $T_1, \dots, T_n$. Since n is large we may further assume that in each T_i there will be at most 1 meteorite impact (a higher number is extremely unlikely). This gives an approximation of X by a binomial random variable $X_n \sim Bin(n, p_n)$ where p_n is unknown. However, since the average total number of impacts[6] in the n-th discretization is np_n, and in the long run it should not depend on n (since we approach X as $n \to \infty$), we arrive at the conclusion that $p_n = \lambda/n$ where λ is the 'intensity parameter' - the average number of impacts per year. By Theorem 2 this will give $X \sim Pois(\lambda)$, and this is indeed what often happens in practice.

[6] We will formalize this in the next lecture