Lecture 1

Probability theory

In this chapter, we recall some basic probability theory that is needed later on. The overview below will be very incomplete, as we will only cover the parts of the theory that we need. For a comprehensive reference, we refer to [Ven13], most of the material covered below can be found in Chapters XV and XVIII if [Ven13].

1.1 Definitions

We start with the definition of a probabilty space. In this definition $\mathcal{P}(A)$ denotes the *power set* of a set A: the set of all subsets of A.

Definition 1.1. A *probability space* is a triple $(\Omega, \Sigma, \mathbb{P})$, where

- Ω is a set,
- $\Sigma \subset \mathcal{P}(\Omega)$ is a σ-algebra and
- $\mathbb{P} : \Sigma \to [0, 1]$ is a probability measure. That is, a measure such that $\mathbb{P}[\Omega] = 1$.

An important example to us will be the following:

Example 1.2. Let Ω be a finite set. We can turn Ω into a probability space by setting $\Sigma = \mathcal{P}(\Omega)$ and

$$\mathbb{P}[A] = \frac{|A|}{|\Omega|} \text{ for all } A \subseteq \Omega,$$

where $|A|$ denotes the cardinality of A. This probability measure is called the *uniform probability measure* on Ω.

From hereon, we will fix the following convention regarding σ-algebras: If Ω is a finite set, then we set $\Sigma = \mathcal{P}(\Omega)$ and if Ω is a topological space then we set $\Sigma = \mathcal{B}(\Omega)$, the Borel algebra of Ω. This convention will allow us not to mention σ-algebras anymore in what follows.

Definition 1.3. Given a probability space $(\Omega, \Sigma, \mathbb{P})$ and a measure space E, an E-valued *random variable* is a measurable function $X : \Omega \to E$. The *expected value* $\mathbb{E}[X]$ of an $\mathbb{R}$-valued random variable X is defined by

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega).$$

The *variance* of X is given by

$$\mathbb{V}\text{ar}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.$$

Two random variables $X, Y : \Omega \to \mathbb{R}$ are called *independent* if

$$\mathbb{P}[X \leq x \text{ and } Y \leq y] = \mathbb{P}[X \leq x] \cdot \mathbb{P}[Y \leq y]$$

for all $x, y \in \mathbb{R}$.

Finally, let $k \in \mathbb{N}$, the quantity

$$\mathbb{E}\left[X^k\right]$$

is called the k^{th} *moment of* X.

Almost all random variables we will consider are real-valued. We remark that in general, the moments of a random variable are not necessarily finite.

Example 1.4. Let $A \subset \Omega$ be measurable. The function $\chi_A : \Omega \to \mathbb{R}$ defined by

$$\chi_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise.} \end{cases}$$

is a random variable that satisfies

$$\mathbb{E}\left[\chi_A\right] = \mathbb{P}\left[A\right].$$

Random variables of this form are called *Bernoulli* random variables.

Often it turns out to be easier to compute expected values than probabilities. In these cases, the following inequality relating the two is useful.

Lemma 1.5. Markov's inequality: *Let $X : \Omega \to \mathbb{R}$ be a random variable such that $\mathbb{E}\left[X\right] < \infty$ and*

$$X(\omega) \geq 0 \ \textit{ for all } \omega \in \Omega$$

Then for all $x \in (0, \infty)$ we have

$$\mathbb{P}\left[X \geq x\right] \leq \mathbb{E}\left[X\right] / x.$$

Proof. We have:

$$x \cdot \mathbb{P}\left[X \geq x\right] = x \cdot \mathbb{E}\left[\chi_{\{\omega \in \Omega;\ X(\omega) \geq x\}}\right] = \mathbb{E}\left[x \cdot \chi_{\{\omega \in \Omega;\ X(\omega) \geq x\}}\right] \leq \mathbb{E}\left[X\right],$$

where the last inequality follows from the fact that

$$x \cdot \chi_{\{\omega \in \Omega;\ X(\omega) \geq x\}}(\omega) \ \leq \ X(\omega)$$

for all $\omega \in \Omega$. □

1.2 The Chen-Stein method

In what follows we will give a self contained account of the Chen-Stein method. Some of the ingredients will however seem to come out of thin air. There is a good motivation for these ingredients, which we will skip in the interest of time.

The goal of this section will be to bound the distance between a random variable X and a Poisson distributed variable. Let us first recall the definition of a Poisson variable.

Definition 1.6. Let $\lambda \in (0, \infty)$. A random variable $X : \Omega \to \mathbb{N}$ is said to be *Poisson distributed with mean* λ if

$$\mathbb{P}[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

for all $k \in \mathbb{N}$.

We also need a notion of distance between random variables:

Definition 1.7. Let E be a measure space with σ-algebra $\mathcal{E}$. Furthermore, let $X, Y : \Omega \to E$ be random variables. The *total variational distance* between X and Y is defined as

$$\mathrm{d_{TV}}(X, Y) = \sup\{|\mathbb{P}[X \in A] - \mathbb{P}[Y \in A]|\,;\ A \in \mathcal{E}\}.$$

If $(X_n)_{n\in\mathbb{N}}$ is a sequence of E-valued random variables so that $\mathrm{d_{TV}}(X_n, X) \to 0$ as $n \to \infty$ then we say that the sequence $(X_n)_{n\in\mathbb{N}}$ *converges to* X *in total variational distance* and write

$$X_n \overset{\mathrm{TV}}{\to} X \ \text{ as } \ n \to \infty.$$

1.2.1 Stein's equation

It turns out that $X : \Omega \to \mathbb{N}$ is Poisson distributed if and only if

$$\mathbb{E}[\lambda g(X+1) - Xg(X)] = 0$$

for all bounded functions $g : \mathbb{N} \to \mathbb{R}$ (See Exercise 1.4 for the easier direction). The basic idea of the Chen-Stein method is that if $\mathbb{E}[\lambda g(X+1) - Xg(X)]$ is close to 0 for all bounded functions g, then X must be close to Poisson distributed.

Given $A \subset \mathbb{N}$, *Stein's equation* is the equation

$$\lambda g(k+1) - kg(k) = \chi_A(k) - \mathbb{E}[\chi_A(Z_\lambda)] \ \text{ for all } k \in \mathbb{N} \tag{1.1}$$

where $\chi_A : \mathbb{N} \to \{0, 1\}$ is defined by

$$\chi_A(n) = \begin{cases} 1 & \text{if } n \in A \\ 0 & \text{otherwise.} \end{cases}$$

We claim that this equation has a unique bounded solution that has value 0 at 0, which we shall denote by $g_A : \mathbb{N} \to \mathbb{R}$. Note that g_A also depends on λ. In order to keep the notation light we shall however suppress λ. The fact that a unique solution g_A exists is clear from the recursive nature of the equation, the proof that it is bounded we shall postpone to Proposition 1.10.

Our earlier remark about using $\mathbb{E}[\lambda g(X+1) - Xg(X)]$ as a measure of the distance to a Poisson variable is made precise by the following theorem.

Theorem 1.8. *Suppose $X : \Omega \to \mathbb{N}$ is a random variable and $Z_\lambda : \Omega \to \mathbb{N}$ is a Poisson distributed random variable with mean $\lambda > 0$. Then*

$$\mathrm{d}_{\mathrm{TV}}(X, Z_\lambda) = \sup \left\{ \left| \mathbb{E}[\lambda g_A(X+1) - X g_A(X)] \right| ;\ A \subset \mathbb{N} \right\}.$$

Proof. Given $A \subset \mathbb{N}$, (1.1) implies that

$$\mathbb{E}[\lambda g_A(X+1) - X g_A(X)] = \mathbb{E}[\chi_A(X)] - \mathbb{E}[\chi_A(Z_\lambda)] = \mathbb{P}[X \in A] - \mathbb{P}[Z_\lambda \in A].$$

As such, filling in the definition of total variational distance leads to the theorem. □

1.2.2 Bounds on Stein's equation

Theorem 1.8 gives us a way to bound the distance between any random variable X and a Poisson variable. Our next and final goal will be to express this bound in terms of more immediate information on X. That is, we will bound $|\mathbb{E}[\lambda g_A(X) - X g_A(X)]|$ in terms of moment(-like expression)s of X.

We start with a bound on solutions of (1.1). Given $A \subset \mathbb{N}$, let us write

$$||g_A|| = \sup_{k \in \mathbb{N}} \{ |g_A(k)| \}.$$

Before we can prove a bound on $||g_A||$, we need the following lemma:

Lemma 1.9. *Let $r, s \in \mathbb{N}$ so that $2r \leq s$. Then:*

$$\sum_{i=0}^{r} \binom{s}{i} \leq \frac{s-r+1}{s-2r+1} \cdot \binom{s}{r}.$$

Proof. Indeed, we have

$$\begin{aligned}\binom{s}{r}^{-1}\sum_{i=0}^{r}\binom{s}{i} &= \sum_{i=0}^{r}\frac{(s-r)!r!}{(s-i)!i!}\\ &= \sum_{i=0}^{r}\frac{(s-r)!r!}{(s-r+i)!(r-i)!}\\ &= \sum_{i=0}^{r}\frac{r(r-1)\cdots(r-i+1)}{(s-r+i)\cdots(s-r+1)}\\ &\leq \sum_{i=0}^{r}\left(\frac{r}{s-r+1}\right)^{i}.\end{aligned}$$

This is a geometric series, so we obtain that

$$\begin{aligned}\binom{s}{r}^{-1}\sum_{i=0}^{r}\binom{s}{i} &\leq \frac{1-\left(\frac{r}{s-r+1}\right)^{r+1}}{1-\frac{r}{s-r+1}}\\ &\leq \frac{s-r+1}{s-2r+1}\end{aligned}$$

for all r, s, where we have used the fact that $r \leq s/2$ for the last step. □

We now have the following bound on $||g_A||$:

Proposition 1.10. *Let $A \subset \mathbb{N}$. Then*

$$||g_A|| \leq 1.$$

Proof. To simplify matters, we define a new function $f : \mathbb{N} \to \mathbb{R}$ by

$$f(k) = \chi_A(k) - \mathbb{E}\left[\chi_A(Z_\lambda)\right]$$

for all $k \in \mathbb{N}$. Note that by definition

$$\mathbb{E}\left[f(Z_\lambda)\right] = 0.$$

Set $g_A(0) = 0$. From (1.1) we obtain that for all $k \in \mathbb{N}$:

$$g_A(k+1) = \frac{1}{\lambda}f(k) + \frac{k}{\lambda}g(k).$$

Hence

$$g_A(k+1) = \frac{1}{\lambda}\sum_{j=0}^{k}\frac{k(k-1)\cdots(k-j+1)}{\lambda^j}f(k-j) = \frac{k!}{\lambda^{k+1}}\sum_{i=0}^{k}\frac{\lambda^i}{i!}f(i).$$

Thus

$$g_A(k+1) = \frac{1}{\lambda\cdot\mathbb{P}\left[Z_\lambda = k\right]}\sum_{j=0}^{k}\mathbb{P}\left[Z_\lambda = j\right]f(j),$$

where $Z_\lambda : \Omega \to \mathbb{N}$ is a Poisson variable with mean λ. Filling in definition of f we get

$$\chi_{[0,k]}(Z_\lambda)f(Z_\lambda) = \chi_{A\cap[0,k]}(Z_\lambda) - \chi_{[0,k]}(Z_\lambda)\mathbb{P}\left[Z_\lambda \in A\right].$$

To shorten notation, let us write:

$$p_\lambda(B) = \mathbb{P}\left[Z_\lambda \in B\right]$$

for all $B \subset \mathbb{N}$ and $U_k = [0,k]\cap\mathbb{N}$. We get

$$\begin{aligned}
\mathbb{E}\left[\chi_{U_k}(Z_\lambda)f(Z_\lambda)\right] &= \mathbb{E}\left[\chi_{A\cap U_k}(Z_\lambda)\right] - \mathbb{E}\left[\chi_{U_k}(Z_\lambda)\right]p_\lambda(A) \\
&= p_\lambda(A\cap U_k) - p_\lambda(U_k)p_\lambda(A) \\
&= p_\lambda(A\cap U_k)\cdot p_\lambda(\mathbb{N}\setminus U_k) - p_\lambda(A\setminus U_k)\cdot p_\lambda(U_k).
\end{aligned}$$

So we obtain

$$g_A(k+1) = \frac{p_\lambda(A\cap U_k)\cdot p_\lambda(\mathbb{N}\setminus U_k) - p_\lambda(A\setminus U_k)\cdot p_\lambda(U_k)}{\lambda\cdot p_\lambda(k)}.$$

Hence

$$\begin{aligned}
|g_A(k+1)| &\leq \frac{\max\left\{p_\lambda(A\cap U_k)\cdot p_\lambda(\mathbb{N}\setminus U_k), p_\lambda(A\setminus U_k)\cdot p_\lambda(U_k)\right\}}{\lambda\cdot p_\lambda(k)} \\
&\leq \frac{p_\lambda(U_k)\cdot p_\lambda(\mathbb{N}\setminus U_k)}{\lambda\cdot p_\lambda(k)}.
\end{aligned}$$

Filling in the Poisson probabilities, we obtain:

$$\begin{aligned}
|g_A(k+1)| &\leq \frac{k!\cdot e^{-\lambda}}{\lambda^{k+1}}\cdot\sum_{i=0}^{k}\frac{\lambda^i}{i!}\sum_{j=k+1}^{\infty}\frac{\lambda^j}{j!} \\
&= k!\cdot e^{-\lambda}\cdot\sum_{i=0}^{k}\frac{\lambda^i}{i!}\sum_{j=0}^{\infty}\frac{\lambda^j}{(j+k+1)!}
\end{aligned}$$

Now we reorder the terms and get:

$$\begin{aligned} |g_A(k+1)| &\le k! \cdot e^{-\lambda} \cdot \sum_{n=0}^{\infty} \lambda^n \sum_{i=0}^{\min\{n,k\}} \frac{1}{i!(n+k+1-i)!} \\ &= k! \cdot e^{-\lambda} \cdot \sum_{n=0}^{\infty} \frac{\lambda^n}{(n+k+1)!} \sum_{i=0}^{\min\{n,k\}} \binom{n+k+1}{i}. \end{aligned}$$

Note that $2 \cdot \min\{n,k\} < n+k+1$ for all $n,k \in \mathbb{N}$, so Lemma 1.9 applies. Hence we get:

$$|g_A(k+1)| \le k! \cdot e^{-\lambda} \cdot \sum_{n=0}^{\infty} \frac{\lambda^n \cdot \binom{n+k+1}{\min\{n,k\}}}{(n+k+1)!} \frac{n+k+2-\min\{n,k\}}{n+k+2-2\cdot\min\{n,k\}}.$$

A straightforward computation shows that

$$\frac{\binom{n+k+1}{n}}{(n+k+1)!} = \frac{1}{n!(k+1)!} \quad \text{and} \quad \frac{\binom{n+k+1}{k}}{(n+k+1)!} = \frac{1}{(n+1)!k!}.$$

This implies that

$$\begin{aligned} |g_A(k+1)| &\le e^{-\lambda} \cdot \left(\sum_{n=0}^{k} \frac{\lambda^n}{n!} \frac{k+2}{(k+1)(k-n+2)} \right. \\ &\quad \left. + \sum_{n=k+1}^{\infty} \frac{\lambda^n}{n!} \frac{n+2}{(n+1)(n-k+2)} \right) \\ &\le e^{-\lambda} \cdot e^{\lambda} \\ &= 1. \end{aligned}$$

□

1.3 Exercises

Exercise 1.1. Let $X : \Omega \to \mathbb{R}$ be a random variable so that

$$\mathbb{E}[X] < \infty \text{ and } \mathbb{E}\left[X^2\right] < \infty.$$

Chebyshev's inequality states that

$$\mathbb{P}\left[|X - \mathbb{E}\left[X\right]| \geq x\right] \leq \frac{\mathbb{V}\mathrm{ar}\left[X\right]}{x^2},$$

for all $x > 0$. Prove Chebyshev's inequality.

Exercise 1.2. Let $X : \Omega \to \mathbb{R}$ be a random variable so that

$$\mathbb{E}\left[X\right] < \infty \text{ and } \mathbb{E}\left[X^2\right] < \infty$$

and

$$X(\omega) \geq 0 \text{ for all } \omega \in \Omega$$

Prove that

$$\mathbb{P}\left[X > 0\right] \geq \frac{\mathbb{E}\left[X\right]^2}{\mathbb{E}\left[X^2\right]}.$$

This inequality is called the *second moment method*.

Exercise 1.3.

(a) Show that two Bernoulli variables $\chi_A, \chi_B : \Omega \to \mathbb{R}$ corresponding to measurable sets $A, B \subseteq \Omega$ are independent if and only if

$$\mathbb{P}\left[A \cap B\right] = \mathbb{P}\left[A\right] \cdot \mathbb{P}\left[B\right].$$

(b) Give an example of a probability space Ω and three random variables $X, Y, Z : \Omega \to \mathbb{R}$ so that all pairs of random variables among $\{X, Y, Z\}$ are independent, but there exists $x, y, z \in \mathbb{R}$ so that

$$\mathbb{P}\left[X \leq x \text{ and } Y \leq y \text{ and } Z \leq z\right] \neq \mathbb{P}\left[X \leq x\right] \cdot \mathbb{P}\left[Y \leq y\right] \cdot \mathbb{P}\left[Z \leq z\right].$$

Exercise 1.4. Show that if a random variable $Z : \Omega \to \mathbb{N}$ is Poisson distributed with mean $\lambda \in (0, \infty)$ then:

$$\mathbb{E}\left[\lambda g(Z+1) - Z g(Z)\right] = 0$$

for all bounded functions $g : \mathbb{N} \to \mathbb{R}$.

Bibliography

[Ven13] Santosh S. Venkatesh. *The theory of probability. Explorations and applications.* Cambridge: Cambridge University Press, 2013.