Probability and Statistics

Solution sheet 3

Solution 3.1 The number of successes in the first group follows the Binomial law $\mathcal{B}(5, 0.5)$, and in the second group follows $\mathcal{B}(5, 0.6)$. The probability to have $k \in \{0, 1, 2, 3, 4, 5\}$ successes in the two cases are respectively

$$\mathbb{P}[X = k] = \binom{5}{k} 2^{-5} \quad \text{for the first group}$$
$$\mathbb{P}[X = k] = \binom{5}{k} 3^k 2^{5-k} 5^{-5} \quad \text{for the second group}$$

	k=0	k=1	k=2	k=3	k=4	k=5
p=0.5	0.0312	0.1562	0.3125	0.3125	0.1562	0.0312
p=0.6	0.0102	0.0768	0.2304	0.3456	0.2592	0.0778

The probability that the first group has at least as many successes as the second group is given by

$$0.03120 * 0.0102 + 0.1562 * (0.0102 + 0.0768) + ... + 0.0312 * 1 = 0.49553028.$$

Solution 3.2

(a) Probability:

(i) With the binomial formula

$$\sum_{k=0}^{n} \mathbb{P}(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$
$$= (p + (1-p))^n = 1.$$

(ii) Consider an urn with replacement containing r red and b blue balls from which we draw a ball n times with replacement, we have that

$$\begin{aligned}\mathbb{P}(\{\text{We have drawn } k \text{ red } n-k \text{ blue}\}) &= \frac{|\{\text{We have drawn } k \text{ red } n-k \text{ blue}\}|}{|\{\text{Possible n-draws}\}|} \\ &= \frac{\binom{n}{k} r^k b^{n-k}}{(r+b)^n} \\ &= \binom{n}{k} \left(\frac{r}{r+b}\right)^k \left(\frac{b}{r+b}\right)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k},\end{aligned}$$

with $p = \frac{r}{r+b}$. Given that in every experiment we draw $k \in \{0, .., n\}$ red balls makes the expression a probability measure, i.e.,

$$
\begin{aligned}
1 &= \mathbb{P}\left(\bigcup_{k=0}^{n}\{\text{We draw } k \text{ red balls}\}\right) \\
&= \sum_{k=1}^{n} \mathbb{P}\left(\{\text{We draw } k \text{ red balls}\}\right) \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}
$$

Expectation:

(i) It holds

$$
\begin{aligned}
\mathbb{E}\left[X\right] &= \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n(n-1)!}{(n-k)!(k-1)!} p^{(k-1)+1} (1-p)^{n-k} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j (1-p)^{(n-1)-j}
\end{aligned}
$$

where we make the change of variables $j = k - 1$. The sum we had is exactly the sum we calculated in the first part for a $Bin(n-1, p)$, so it is 1. Thus:

$$\mathbb{E}\left[X\right] = np.$$

(ii) We know that the amount of red balls that are taken out at time n in a experiment with replacement have the distribution of X. So

$$
\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}\left[\sum_{j=1}^{n} \mathbb{1}_{\{\text{In the } j\text{-th draw we get a red ball}\}}\right] \\
&= \sum_{j=1}^{n} \mathbb{P}\left(\{\text{In } j\text{-th draw we get a red ball}\}\right).
\end{aligned}
$$

The probability that in the j-th draw we get a red ball is p, so:

$$\mathbb{E}\left[X\right] = np.$$

(b) Probability:

(i) Note that

$$\begin{aligned}\sum_{n=0}^{N}\binom{N}{n}x^n &= (1+x)^N \\ &= (1+x)^K(1+x)^{N-K} \\ &= \sum_{k=0}^{K}\binom{K}{k}x^k\sum_{j=0}^{N-K}\binom{N-K}{j}x^j \\ &= \sum_{k=0}^{K}\sum_{j=0}^{N-K}\binom{K}{k}\binom{N-K}{j}x^{k+j} \\ &= \sum_{n=0}^{N}\sum_{k=\max\{0,K+n-N\}}^{\min\{n,K\}}\binom{K}{k}\binom{N-K}{n-k}x^n,\end{aligned}$$

where we made the change of variables $n=k+j$. Given that two polynomials are equal if and only if all of its coefficients are equal we have that

$$\sum_{k=\max\{0,K+n-N\}}^{\min\{n,K\}}\binom{K}{k}\binom{N-k}{n-k}=\binom{N}{n}$$

which implies

$$\sum_{k=\max\{0,K+n-N\}}^{\min\{n,K\}}\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}=1,$$

so the definition gives a probability measure.

(ii) If you have N balls, K of which are red and $N-K$ blue and you are drawing them out without replacement. We have that the event $B :=$ "after the n-th draw we have extracted k red balls and $n-k$ blue balls" is given by

$$\begin{aligned}\mathbb{P}(B) &= \frac{|\{\text{Ways of taking out } k \text{ red balls and } n-k \text{ blue balls }\}|}{|\{\text{Ways of taking out n balls}\}|} \\ &= \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.\end{aligned}$$

Given that in every experiment we extract $k\in\{0,..,n\}$ red balls, we obtain a probability measure.

Expectation:

(i) We compute

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=\max\{1,n+K-N\}}^{\min\{n,K\}} k\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \\
&= K\sum_{k=\max\{1,n+K-N\}}^{\min\{n,K\}} \frac{\binom{K-1}{k-1}\binom{(N-1)-(K-1)}{(n-1)-(k-1)}}{\binom{N}{n}} \\
&= K\frac{1}{\binom{N}{n}}\sum_{u=\max\{0,n+K-N-1\}}^{\min\{n-1,K-1\}} \binom{K-1}{u}\binom{(N-1)-(K-1)}{(n-1)-u} \\
&= K\frac{\binom{N-1}{n-1}}{\binom{N}{n}} = n\frac{K}{N}.
\end{aligned}$$

where we have used the sum we calculated in the first part for a $Hyp(N-1, K-1, n-1)$.

(ii) We see that $X = \sum_{j=1}^{n} \mathbb{1}_{B_j}$ where B_j is "in the n-th drawing we take a red ball". We also have

$$\mathbb{P}(B_j) = \mathbb{P}(B_1) = K/N,$$

and since $\sum_{j=1}^{n} \mathbb{1}_{B_j} = X$ we conclude

$$\mathbb{E}(X) = \mathbb{E}\left[\sum_{j=1}^{n} \mathbb{1}_{B_j}\right] = n\mathbb{P}(B_1) = n\frac{K}{N}.$$

Solution 3.3

(a) The event "the first client comes later than t hours" is the same as "nobody comes during the first t hours". The number of visits during the first t hours has the Poisson distribution with mean t, hence nobody comes with probability e^{-t}. The arrival time of the first client follows the exponential law of parameter 1.

(b) Similarly, the number of visits during the first hour has the law Poi(1). The probability of having strictly more than two clients is given by

$$1 - e^{-1}\left(\frac{1^0}{0!} + \frac{1^1}{1!} + \frac{1^2}{2!}\right) = 1 - 5/(2e) \approx 0.0803.$$

(c) This event is included in "at least one client comes in the time interval $[t - \varepsilon/2, t + \varepsilon/2]$", which has probability

$$1 - e^{-\varepsilon} \sim \varepsilon$$

as $\varepsilon \to 0$. Thus the probability that a client comes at the exact time t is 0.

Although "a client comes during the first hour" is the union over all $t \leq 1$ of "a client comes at time t", the former event has a positive probability as we have seen in the previous question. It is not contradictory because its probability is the sum of the probabilities of an uncountable family of events.

(d) We show that the sum X of two independent Poisson random variable of parameter a and b

is distributed as a Poisson$(a+b)$. Indeed, by independence:

$$\begin{aligned}
\mathbb{P}[X=k] &= \mathbb{P}\left[\bigcup_{i=0}^{k}\{X_a=i, X_b=k-i\}\right] \\
&= \sum_{i=0}^{k}\mathbb{P}[\{X_a=i, X_b=k-i\}] \\
&= \sum_{i=0}^{k}\mathbb{P}[\{X_a=i\}]\,\mathbb{P}[\{X_b=k-i\}] \\
&= \sum_{i=0}^{k} e^{-a-b}\frac{a^i}{i!}\frac{b^{k-i}}{(k-i)!} \\
&= e^{-a-b}\frac{b^k}{k!}\sum_{i=0}^{k}\left(\frac{a}{b}\right)^i\frac{k!}{i!(k-i)!} \\
&= e^{-a-b}\frac{b^k}{k!}\left(1+\frac{a}{b}\right)^k \\
&= e^{-a-b}\frac{(a+b)^k}{k!}.
\end{aligned}$$

Thus the total number of clients coming to both stores during time interval of length t follows a Poisson law with parameter $(\lambda+\mu)t$, which means that the total arrival of clients is a Poisson process with rate $\lambda+\mu$.

Solution 3.4

(a) Clearly $2^{\mathbb{N}}$ is a σ-algebra. $\mathbb{P}$ is a probability measure because:

$$\mathbb{P}(\{i\}) \geq 0,$$

and for $W_i \subseteq \mathbb{N}$, $i \in \mathbb{N}$, disjoint sets,

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{\infty} W_i\right) &= \sum_{n\in\bigcup_{i=1}^{\infty} W_i} f(n) \\
&= \sum_{i=1}^{\infty}\sum_{n\in W_i} f(n) \\
&= \sum_{i=1}^{\infty}\mathbb{P}(W_i).
\end{aligned}$$

Additionally it is clear from the definition that $\mathbb{P}(\mathbb{N})=1$. Then $(\mathbb{N}, 2^{\mathbb{N}}, \mathbb{P})$ is a probability space.

(b) Let p be a prime number, then $N_p = \{np \mid n \in \mathbb{N}\}$. We have

$$\begin{aligned}
\mathbb{P}(N_p) &= \frac{1}{\zeta(s)}\sum_{n=1}^{\infty}\frac{1}{(np)^s} \\
&= \frac{1}{p^s}\left(\frac{1}{\zeta(s)}\sum_{n=1}^{\infty}\frac{1}{n^s}\right) \\
&= \frac{1}{p^s}.
\end{aligned}$$

(c) Take $(p_{i_k})_{k=1}^m$ a finite family of prime numbers. Then $\bigcap_{k=1}^m N_{p_{i_k}} = \{n \in \mathbb{N} : n \prod_{k=1}^m p_{i_k}\}$, so

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{k=1}^m N_{p_{i_k}}\right) &= \frac{1}{\zeta(s)} \sum_{n\in\mathbb{N}} \frac{1}{n^s \prod_{k=1}^m p_{i_k}^s} \\
&= \frac{1}{\prod_{k=1}^m p_{i_k}^s} \frac{1}{\zeta(s)} \sum_{n\in\mathbb{N}} \frac{1}{n^s} \\
&= \prod_{k=1}^m \frac{1}{p_{i_k}^s} \\
&= \prod_{k=1}^m \mathbb{P}(N_{p_{i_k}}),
\end{aligned}$$

and the $N_{p_{i_k}}$'s are independent.

(d) Note that $\bigcap_{k=1}^\infty N_{p_k}^c = \{1\}$. Then, with question (c).

$$\begin{aligned}
\frac{1}{\zeta(s)} &= \mathbb{P}\left(\bigcap_{k=1}^\infty N_{p_k}^c\right) \\
&= \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{k=1}^m N_{p_k}^c\right) \\
&= \lim_{m\to\infty} \prod_{k=1}^m (1 - p_k^{-s}) \\
&= \prod_{k=1}^\infty (1 - p_k^{-s}),
\end{aligned}$$

Thus,

$$\zeta(s) = \frac{1}{\prod_{k=1}^\infty (1 - p_k^{-s})}.$$