TD-Probability

Marco Biroli

October 26, 2020

Chapter 1

TD 1

1.1 A strategic choice.

Let $X \in \{0,1\}^3$ (resp. Y) be the random variable corresponding to the results of the matches using the first strategy (resp. the second strategy). Then we have that (let $D = \{(1,1,1),(1,1,0),(0,1,1)\}$):

$$P(X \in D) = a^2b + ab(1-a) + (1-a)ba = ab(2-a)$$

Similarly:

$$P(Y \in D) = b^2a + ba(1-b) + (1-b)ab = ba(2-b)$$

Then since $a > b$ we have that $P(X \in D) < P(Y \in D)$, hence the winning strategy is BAB.

1.2 Derangements

1.2.1

Let E be a finite set and $A, B \subseteq E$. We denote by 1_A the indicator function of A and $\bar{A}$ the complement of A. Then we have that:

$$1_{\bar{A}} = 1 - 1_A \text{ and } 1_{A\cap B} = 1_A \cdot 1_B \text{ and } 1_{A\cup B} = 1_A + 1_B - 1_{A\cap B}$$

1.2.2

We will prove this by induction on n. The base case $n = 1$ as well as $n = 2$ are trivially satisfied. Now assume that this is satisfied for n then we have that (using the induction hypothesis for $n = 2$):

$$\text{card}\left(\bigcup_{i=1}^{n} A_i \bigcup A_{n+1}\right) = \text{card}\left(\bigcup_{i=1}^{n} A_i\right) + \text{card}(A_{n+1}) - \text{card}\left(\left(\bigcup_{i=1}^{n} A_i\right)\bigcap A_{n+1}\right)$$

Now we develop the last term into:

$$\left(\bigcup_{i=1}^{n} A_i\right)\bigcap A_{n+1} = \bigcup_{i=1}^{n}\left(A_i \bigcap A_{n+1}\right)$$

Now applying the induction hypothesis gives the desired result.

1.2.3

Let A_i be the set of permutations that fixes point i. Then from the inclusion-exclusion principle we have:

$$D_n = n! - \text{card}\left(\bigcup_{i=1}^{n} A_i\right) = n! - \sum_{k=1}^{n}(-1)^{k-1}\binom{n}{k}(n-k)! = n!\sum_{k=2}^{n}\frac{(-1)^k}{k!}$$

1.2.4

The probability that no one gets their jacket corresponds to the probability of having a derangement in other words:

$$p_n = \frac{D_n}{n!} = \sum_{k=2}^{n}\frac{(-1)^k}{k!} \overset{n\to\infty}{\longrightarrow} \frac{1}{e}$$

1.2.5

We have that:

$$D_{n,l} = \binom{n}{l} D_{n-l} = \binom{n}{l}(n-l)!\sum_{k=2}^{n-l}\frac{(-1)^k}{k!} = \frac{n!}{l!}\sum_{k=2}^{n-l}\frac{(-1)^k}{k!}$$

Hence the probability that eaxctly l people leave with their jackets is:

$$p_l = \frac{D_{n,l}}{n!} = \frac{1}{l!}\sum_{k=2}^{n-l}\frac{(-1)^k}{k!}$$

1.2.6

The probability that a given person gets back their jacket is $p_s = \frac{1}{n}$. The probability that at least one person gets back their jacket is:

$$p_a = 1 - p_n = 1 - \sum_{k=2}^{n}\frac{(-1)^k}{k!}$$

Notice that $p_s < p_a$.

1.3 Balls in bins

1.3.1

(a) If all the balls are distinguishable then we have $\Omega = [\![1, n]\!]^r$ is the set of tuples where each element corresponds to where the i-th ball has been sent to. Then $\mathcal{F} = \mathcal{P}(\Omega)$ and since each event is sampled uniformly at random we have that:

$$\forall \omega \in \Omega, P(\omega) = \frac{1}{|\Omega|} = \frac{1}{n^r}$$

Then the probability of $(r_1, \cdots, r_n)$ is given by:

$$P[(r_1, \cdots, r_n)] = P[\{\omega \in \Omega : \forall i \in [\![1, n]\!] \,\#\{b \in \Omega : b = i\} = r_i\}] = \frac{1}{n^r}\binom{r}{r_1, r_2, \ldots, r_n}$$

(b) Now we have that $\Omega = \{(r_i \in \mathbb{N} : i \in [\![1, n]\!]) : \sum_{i=1}^n r_i = r\}$. Again we have that $\mathcal{F} = \mathcal{P}(\Omega)$. Then we have that:

$$P[(r_1, \ldots, r_n)] = \frac{1}{|\Omega|} = \frac{1}{\binom{r+n-1}{n-1}}$$

(c) Now we have that $\Omega = \{s \in \{0,1\}^n : \sum_{i=1}^n s_i = r\}$ corresponding to the tuple indicating if each state is occupied or not. Once again $\mathcal{F} = \mathcal{P}(\Omega)$. Now the probability is given by:

$$P[(r_1, \ldots, r_n)] = \frac{1}{|\Omega|} = \frac{1}{\binom{n}{r}}$$

1.3.2

The probability that at least two have the same birthday is the 1 minus the probability that none of them share a birthday. The probability that none of them share a birthday is given by $\frac{r!}{n^r}\binom{n}{r}$. Hence the probability that at least two people share a birthday is given by: $1 - \frac{r!}{n^r}\binom{n}{r}$.

1.3.3

Days are bins, accidents are distinguishable balls hence the probability is given by:

$$\frac{\binom{r}{n} n^{r-n}}{n^r} = n^{-n}\binom{r}{n}$$

1.3.4

Chapter 2

TD2

2.1 Symmetric Random Walk

Consider a balanced coin drawn n times. Denote $X_1, \cdots, X_n$ the results and S_k the partial sums.

2.1.1

The law of S_k is given by:

$$p_{n,r} = P(S_n = r) = \frac{1}{2^n}\binom{n}{\frac{n+r}{2}}$$

2.1.2

The number of paths from $(0,0)$ to $(2n+2,0)$ never zero are equal to the number of paths from $(1,1)$ to $(2n+1,1)$ which always stay above or equal to the line $y=1$. Hence rescaling the y-axis by a factor 1 we get a bijection in between the strictly positive walks from $(0,0)$ to $(2n+2,0)$ with the positive or zero walks from $(0,0)$ to $(2n,0)$. Hence for symmetric random walks the number of random walks going from $(0,0)$ to $(2n+2,0)$ never touching the axis is twice as much as the number of walks from $(0,0)$ to $(2n,0)$ being always positive or 0. Furhtemore there are 4 times more walks going from 0 to $2n+2$ which therefore gives the desired result.

2.1.3

We now that the end of the random walk is going to be given by $a-b$. Now the number of possible only positive walks is given by the number of walks from $(1,1)$ to $(a+b, a-b)$ minus the number of walks from $(1,-1)$ to $(a+b, a-b)$ by the reflexion principle. Hence we get that:

$$p = p_{a+b-1,a-b-1} - p_{a+b-1,a-b+1} = \frac{1}{2^n}\frac{a-b}{a+b}\binom{a+b}{a}$$

2.1.4

a)

Up to a re-scaling of the y-axis we have the equivalent problem of computing the number of paths that go from $(0,-r)$ to $(n, k-r)$ and which touch the x-axis at least once. Now notice that from the reflexion principle this is equal to the number of paths from $(0,r)$ to $(n, k-r)$ and up to a second shifting this is equal to the number of paths from $(0,0)$ to $(n, k-2r)$. Hence the desired probability is given by $p_{n,k-2r} = p_{n,2r-k}$.

b)

Then for any r we have that:

$$\begin{aligned}
P(\max\{S_1,\cdots,S_n\}=r) &= \sum_{k=-\infty}^{+\infty} P(S_n=k,\max\{S_1,\cdots,S_n\}=r) \\
&= \sum_{k=-\infty}^{+\infty} \left(P(S_n=k,\max\{S_1,\cdots,S_n\}\geq r) - P(S_n=k,\max\{S_1,\cdots,S_n\}>r)\right) \\
&= \sum_{k=-\infty}^{+\infty} \left(P(S_n=k,\max\{S_1,\cdots,S_n\}\geq r) - P(S_n=k,\max\{S_1,\cdots,S_n\}\geq r+1)\right) \\
&= \sum_{k=-\infty}^{+\infty} (p_{n,k-2r} - p_{n,k-2r-2}) = p_{n,r} + p_{n,r+1}
\end{aligned}$$

c)

This can be re-written as:

$$P(S_n=0,S_1<0,\cdots,S_{n-1}<0,S_0=-r) = P(S_n=-r,S_1<0,\cdots,S_{n-1}<0,S_0=0)$$

Now again notice that this corresponds to a symmetric re-writing of the problem 3. Hence we get immediately that:

$$\hat{p}_{n,r} = P(S_n=r,S_1<r,\cdots,S_{n-1}<r) = \frac{1}{2^n}\frac{r}{n}\binom{n}{\frac{n+r}{2}} = \frac{r}{n}p_{n,r}$$

d)

2.2 Geometric and negative-binomial laws.

2.2.1

We want to compute:

$$P(T_1 = t+1)$$

For such a thing to be the case we need to have thrown tails successively t times and heads the last time. Hence the probability is given by:

$$P(T_1-1=t) = P(T_1=t+1) = (1-p)^t p = \mathcal{G}(p)$$

The expectancy of $\mathcal{G}(p)$ is given by:

$$\mathbb{E}[\mathcal{G}(p)] = \sum_{t=0}^{+\infty}(1-p)^t pt = p\cdot\frac{1-p}{p^2} = \frac{1-p}{p}$$

The variance of $\mathcal{G}(p)$ is given by:

$$\mathrm{Var}[\mathcal{G}(p)]^2 = \frac{p^2-3p+2}{p^2} - \frac{1-2p+p^2}{p^2} = \frac{1-p}{p^2}$$

2.2.2

A geometric law corresponds to something not happening t times and then happening at the $t+1$ time. Then the infimum of two geometric laws corresponds to two things not happening t times and one of them happening at the $t+1$ time. The probability of which is given as follows:

$$P[(\inf(S_1,S_2)=s] = (1-p)^{2(s-1)}(p^2+2p(1-p)) = (1-p)^{2(s-1)}(2p-p^2) = (1-p)^{2(s-1)}(1-(1-p)^2)$$

Hence the infimum is a geometric variable with law $\mathcal{G}(1-(1-p)^2)$

2.2.3

We want to compute $P(T_m - m = k)$. The number of possible outcomes for which $T_m - m = k$ is given by $\binom{k+m-1}{m-1}p^m(1-p)^k = Neg(m,p)$. Now using the formula we get that:

$$\sum_{k\geq 0} Neg(m,p)[k] = p^m(1-(1-p))^{-m} = 1$$

2.2.4

Notice that we can re-write $T_m - m$ as the number of steps before the first H plus the number of steps between the first and second head etc. now the number of steps in between two succesive heads is given by $\mathcal{G}(p)$ then the result follows.

2.2.5

Then we have that:

$$\mathbb{E}[T_m - m] = m\mathbb{E}[\mathcal{G}(p)] = m\frac{1-p}{p}$$

Hence:

$$\mathbb{E}[T_m] = \frac{m}{p}$$

Similarly we get:

$$Var(T_m) = m\frac{1-p}{p^2}$$

2.3 Thinning and Poisson random variables

We have that:

$$P[Y = k|X = n] = \binom{n}{k}p^k(1-p)^{n-k}$$

Then from the law of total probability we have that:

$$P[Y = k] = \sum_{n=k}^{+\infty}\binom{n}{k}p^k(1-p)^{n-k}\frac{\lambda^n e^{-\lambda}}{n!} = \frac{p^k e^{-\lambda}\lambda^k}{k!}\sum_{n=0}^{+\infty}(1-p)^n\frac{\lambda^n}{(n)!} = \frac{(\lambda p)^k e^{-\lambda p}}{k!}$$

2.4 Conditional Probabilities

1. (a) $P(B|A) = \frac{P(A\cap B)}{P(B)} = \frac{P(A)}{P(B)}\frac{P(A\cap B)}{P(A)} = \frac{P(A)}{P(B)}P(A|B)$.

   (b) From the previous question we have:

   $$P(B|A) = \frac{P(A)}{P(B)}P(A|B) = \frac{P(A|B)}{P(B)}\sum_{i\in\mathbb{N}}P(A|C_i)P(C_i)$$

   (c) Let X be the r.v. corresponding to the number of children of law p_i and A the event "has no daughter". Then:

   $$P(X = 1|A) = \frac{P(A|X = 1)}{P(X = 1)}\sum_{i\in\mathbb{N}}P(A|X = i)P(X = i) = \frac{1}{2p_1}\sum_{i\in\mathbb{N}}\frac{p_i}{2^i}$$

2. The law of (X, Y) is given by:

   $$P((X,Y) = (x,y)) = \frac{1}{6}\cdot\frac{\delta_{y\leqslant x}}{x} \quad \text{for} \quad (x,y)\in [\![1,6]\!]^2$$

   Then:

   $$P(X = x) = \sum_{y=1}^{6}P((X,Y) = (x,y)) = \sum_{y=1}^{x}\frac{1}{6x} = \frac{1}{6}$$

   Similarly:

   $$P(Y = y) = \sum_{x=1}^{6}P((X,Y) = (x,y)) = \frac{1}{6}\sum_{x=y}^{6}\frac{1}{x}$$

2.5 Change of variables.

1. Let U be a r.v. with uniform law over $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Then:

$$P(\tan(U) \leq x) = P(U \leq \tan^{-1}(x)) = \frac{\arctan(x) - \frac{\pi}{2}}{\pi}$$

Then the pdf of $\tan U$ is given by:

$$f(x) = \frac{1}{\pi(1+x^2)}$$

Then the value of $\mathbb{E}[|\tan U|]$ is given by:

$$\mathbb{E}[|\tan U|] = \int_{-\infty}^{+\infty} \frac{|x|}{\pi(1+x^2)} \mathrm{d}x$$

Which diverges.

2. We have that $X = \cos\theta$ and $Y = \sin\theta$ and θ is a r.v. with uniform law on $[0, 2\pi[$. Then:

$$f_X(x) = f_{\cos\theta}(x) = \frac{1}{\pi}\left|\frac{\mathrm{d}}{\mathrm{d}x}\arccos x\right| = \frac{1}{\pi\sqrt{1-x^2}}$$

By symmetry it is immediate that $P(X = x) = P(Y = x)$. Then $z = X + Y = \cos\theta + \sin\theta = \sqrt{2}\cos\big(\theta - \frac{\pi}{4}\big)$. Then:

$$f_Z(z) = f_x\left(\frac{z}{\sqrt{2}}\right)\frac{1}{\sqrt{2}} = \frac{1}{\pi}\frac{1}{\sqrt{2-z^2}}$$

2.6 Random Variables

1. Let $X(\omega) = a1_A(\omega) + b1_B(\omega)$ and $C \in \mathcal{B}(\mathbb{R})$. Then:

$$\sigma(X) = \langle \emptyset, A, B, A\cup B, A\cap B\rangle \quad \text{and} \quad P(X = x) = \begin{cases} P(A) & \text{if } x = a \\ P(B) & \text{if } x = b \\ P(A\cap B) & \text{if } x = a+b \end{cases}$$

2. Then:

$$P_X = \frac{1}{2}\delta_1 + \frac{1}{2}Unif([0,1]) \quad \text{and} \quad \sigma(X) = \langle \emptyset, \mathcal{B}([0,1/2]), [1/2,1]\rangle$$

3. Then:

$$P(X \leq x) = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{\mathrm{d}x}{2} = \sqrt{x} \quad \text{and} \quad \sigma(X) = \{A \in \mathcal{B}([-1,1]) : A = -A\}$$

Chapter 3

TD3

3.1 Gamma Law

1. We have that:

$$f_X(x) = f_{Z/\lambda}(x) = h_{a,1}(\lambda x)\lambda = 1_{x\geq 0}\frac{1}{\Gamma(a)}1^a(\lambda x)^{a-1}\exp(-\lambda x)\lambda = h_{a,\lambda}(x)$$

2. We have that:

$$E[Z] = \int_0^{+\infty}\frac{x^{a-1}e^{-x}}{\Gamma(a)}x\mathrm{d}x = \frac{1}{\Gamma(a)}\int_0^{+\infty}x^{(a+1)-1}e^{-x}\mathrm{d}x = \frac{\Gamma(a+1)}{\Gamma(a)} = a+1$$

Which also immediately gives:

$$E[X] = E[Z/\lambda] = \frac{a}{\lambda}$$

Then the variance is given by a similar integration which yields:

$$Var[Z] = \frac{\Gamma(a+2)}{\Gamma(a)} - \frac{\Gamma(a+1)}{\Gamma(a)} = a$$

Which also gives:

$$Var[X] = Var[Z/\lambda] = \frac{a}{\lambda^2}$$

3. Using question 1 it is sufficient to show the case where X, Y have laws $\mathcal{G}(a,1)$ and $\mathcal{G}(b,1)$. Then:

$$f_Z(z) = f_X \star f_Y(z) = \int f_{X,Y}(x, z-x)\mathrm{d}x = \int f_X(x)f_Y(z-x)\mathrm{d}x$$

Now replacing with the laws we get:

$$f_Z(z) = \int_0^z \frac{1}{\Gamma(a)\Gamma(b)}x^{a-1}(z-x)^{b-1}e^{-\lambda x-\lambda(z-x)}\mathrm{d}x = \frac{z^{a-1+b-1+1}e^{-\lambda z}}{\Gamma(a)\Gamma(b)}\int_0^1 u^{a-1}(1-u)^{b-1} = h_{a+b,1}(z)$$

Where the last equality follows from normalization. For a proof check the next question.

3.2 Beta law

1. We compute:

$$\int_0^1 x^{a-1}(1-x)^{b-1}\mathrm{d}x =$$

2. We introduce $V = X$ and make the change of variable $(X,Y) \to (Z,V)$. Which explicitly gives $(Z,V) = (XY, X)$ Then the Jacobian determinant is given by:

$$\begin{vmatrix} y & x \\ 1 & 0 \end{vmatrix} = -x = -v$$

Then we have:

$$f_{Z,V}(z,v) = f_{x,y}(v, \frac{z}{v})\frac{1}{|v|}$$

And integrating gives:

$$f_Z(z) = \int f_{X,Y}(x, \frac{z}{x})\frac{1}{|x|}\mathrm{d}x = \int f_X(x) f_Y(\frac{z}{x})\frac{1}{|x|}\mathrm{d}x$$

Then plugging in the laws gives the desired result.

3.3 Gaussian Law

1. We have:

$$E[X] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \exp\left[-\frac{x^2}{2}\right]x\mathrm{d}x = 0$$

Since the integrand is odd. Then the variance is given by:

$$E[X^2] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} \exp\left[-\frac{x^2}{2}\right]x^2\mathrm{d}x = 1$$

2. The law of $Y = m + \sigma X$ is given by:

$$f_Y(y) = f_X(\frac{y-m}{\sigma})\frac{1}{\sigma} = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(\frac{y-m}{\sigma})^2}{2}\right] = h_{m,\sigma^2}(y)$$

Then we have that:

$$E[Y] = m \text{ and } Var[Y] = \sigma^2$$

3. Let $Z = X/Y$ and $V = Y$ then the Jacobian determinant is given by:

$$\begin{vmatrix} \frac{1}{Y} & -\frac{X}{Y^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{Y}$$

Then we have that:

$$f_{Z,V}(z,v) = f_{X,Y}(zv, v)|y| \Rightarrow f_Z(z) = \int f_X(zv) f_Y(v)|v|\mathrm{d}v$$

Which when replacing with the laws gives:

$$f_Z(z) = \frac{1}{\pi}\frac{1}{1+z^2}$$

From symmetry the law of $1/Z$ is identical.

Chapter 4

TD 4

4.1 Around the notion of independence.

1. A and B are independent, B and C also and A and C also. However A, B, C is not independent.

2. $P(A) = P(A \cap A) = P(A)P(A) = P(A)^2$ then $P(A) = 0$ or $P(A) = 1$. Similarly $P(X = a) = P(X = a|X = b) = \delta_{ab}$. So X must be single valued and have that value with probability 1.

3. We have that $E[X] = E[Y] = 0$. Since X, Y are independent we also have that $E[Z] = E[XY] = E[X]E[Y] = 0$. Hence we have that:

$$Cov[X, Z] = E[(X - E[X])(Z - E[Z])] = E[XZ] = E[X^2Y] = E[X^2]E[Y] = E[X^2] \cdot 0 = 0$$

X and Z are not independent because:

$$P(Z = 0|X = 0) = 1 \neq P(Z = 0)$$

4.2 Convergence in probability of random variables.

1. We have that:

$$\forall \varepsilon > 0, \lim_{n \to +\infty} P(|X_n - X| > \varepsilon) = 0$$

Now notice that:

$$P(|YX_n - YX| > \varepsilon) = P(|Y||X_n - X| > \varepsilon)$$

Since this must be true for all ε it is equivalent to saying that:

$$P(|X_n - X| > \varepsilon \wedge |Y| > \varepsilon) \leq P(|X_n - X|)$$

Which by hypothesis converges to 0.

2. Fix ε then:

4.3 Different modes of convergence of random variables.

1. (X_n) converges to δ_0 almost surely from construction. Then we have that:

$$E[|X_n - \delta_0|] = E[|X_n|] = 0 \cdot (1 - \frac{1}{n^2}) + n^2 \cdot \frac{1}{n^2} = 1$$

Hence it does not converge to δ_0 in L^1.

2. (a) We have that:

$$E[X_n] = 0(1 - \frac{1}{n}) + 1 \cdot \frac{1}{n} = \frac{1}{n} \overset{n \to +\infty}{\longrightarrow} 0$$

Hence X_n goes to 0 in L^1 and therefore also in probability.

(b) We have that:

$$P(\limsup A_n) = P(\bigcap_m \bigcup_{n \geq m} A_n) = P(\lim_{n \to +\infty} \exists m \geq n, X_m = 1)$$

Chapter 5

5.1 Cauchy random variables

1. We pass by the characteristic function:
$$\varphi_X(k) = E[e^{ikX}] = \int_{-\infty}^{+\infty} \frac{ae^{ikx}}{\pi(x^2+a^2)}\mathrm{d}x$$
Then from the residue theorem we get that:
$$\varphi_X(k) = e^{-a|k|}$$
Then:
$$\varphi_{X+Y}(k) = E[e^{ikX}e^{ikY}) = E[e^{ikX}]E[e^{ikY}] = \varphi_X(k)\varphi_Y(k) = e^{-(a+b)|k|}$$
Hence $X+Y \sim \text{Cauchy}(a+b)$. Then:
$$f_{X/\lambda}(u) = f_X(\lambda u)\lambda = \frac{a}{\pi(\lambda^2u^2+a^2)}\lambda = \frac{a/\lambda}{\pi(u^2+a^2/\lambda^2)} = \text{Cauchy}(a/\lambda)$$

2. Generalizing the previous result we have that $S_n \sim \text{Cauchy}(na)$ and hence $S_n/n \sim \text{Cauchy}(a)$. Hence it does not converge to a constant. There is no contradiction because the strong law requires that the expectancy is defined, which is not the case of Cauchy random variables.

5.2 Almost sure convergence

From the B.C lemma we know that $P(|X_{n+1}-X_n| > a_n \;\text{ i.o.}) = 0$. Hence we have that $P(|X_{n+1}-X_n| \leq a_n \;\text{ i.o}) = 1$. Now since the series $\sum_n a_n < \infty$ we must have that $a_n \to 0$ and hence $(X_n(\omega))_{n\geq 1}$ is a Cauchy sequence and since it is real valued and the reals are complete is converges.

5.3 A version of the strong law of large numbers

1. We have that:
$$E(|\frac{S_n}{n}|^4) = E((\frac{S_n}{n})^4) = \frac{1}{n^4}E((\sum_{i=1}^n X_i)^4) = \frac{1}{n^4}(E(\sum_{i=1}^n X_i^4)+E(\sum_{i,j=1}^n X_i^2X_j^2)+E(\sum_{i,j,k} X_i^2X_jX_k)+E(\sum_{i,j,k,\ell} X_iX_jX_kX_\ell))$$
Which simplifies to:
$$\frac{1}{n^4}(nE[X^4]+n(n-1)(4E(X^3)E(X)+3E(X^2)E(X^2))+6n(n-1)(n-2)E(X^2)E(X)^2+n(n-1)(n-2)(n-3)E(X)^4)$$
Now using the fact that $E[X]=0$ hence it simplifies to:
$$\frac{1}{n^2}((3(1-\frac{1}{n})E[X^2]^2+\frac{1}{n^3}E[X^4])$$

2. We have from the Chebyshev inequality that:
$$P(|\frac{S_n}{n}| \geq \varepsilon) = P(|\frac{S_n}{n}|^4 \geq \varepsilon^4) \leq \frac{1}{\varepsilon^4}E[|\frac{S_n}{n}|^4] = a_n$$
Then notice that:
$$\sum_n a_n = \frac{1}{\varepsilon^4}\sum_n E[|\frac{S_n}{n}|^4] < +\infty$$
From the previous question since every term converges.

3. From the dominated convergence theorem we can apply B.C lemma and we know that only finitely often will the sequence diverge. Hence it will almost surely converge.

4. Up to centering the variables this follows immediately.

5.4 Bernstein Polynomials

1. We know that $S_n \sim \text{Binomial}(n, x)$ then $E[S_n] = nx$. Then we have:

$$E[f(S_n/n)] = \sum_{k=0}^{n} f(\frac{k}{n}) \binom{n}{k} x^k (1-x)^{n-k} = P_n(x)$$

2. Since $E[|S_n/n|] < +\infty$ we can apply the strong law of large numbers and we know that $S_n/n \to x$ almost surely and consequently we know that $f(S_n/n) \to f(x)$ almost surely. Hence we have that $E[f(S_n/n)] \to E[f(x)] = f(x)$ almost surely and hence $P_n(x) \to f(x)$ for every x almost surely.

3. The variance of S_n is given by $\text{Var}(S_n) = nx(1-x) \leq \frac{n}{4}$.

4. We have that:

$$P(|S_n - nx| \geq n\varepsilon) \leq \frac{\text{Var}(S_n)}{n^2\varepsilon^2} = \frac{1}{4n\varepsilon}$$

Chapter 6

6.1 Characteristic functions

1. The law of a Poisson random variable is given by:

$$P(X = k) = \frac{a^k e^{-a}}{k!}$$

Then the characteristic function is given by:

$$\varphi_X(k) = E(e^{ikX}) = \sum_{n=0}^{+\infty} e^{ikn} \frac{a^n e^{-a}}{n!} = e^{-a} \sum_{n=0}^{+\infty} \frac{(e^{ik}a)^n}{n!} = e^{-a} e^{ae^{ik}} = e^{a(e^{ik}-1)}$$

2. Let $Y \sim \text{Poisson}(b)$. Then:

$$\varphi_{X+Y}(k) = E(e^{ik(X+Y)}) = E(e^{ikX})E(e^{ikY}) = e^{a(e^{ik}-1)} e^{b(e^{ik}-1)} = e^{(a+b)(e^{ik}-1)}$$

Hence $X + Y \sim \text{Poisson}(a + b)$.

6.2 A distributional equation.

1. We have that:

$$\varphi_X(k) = \int_S e^{ikx} P_X(x) \mathrm{d}x \quad \text{then} \quad |\varphi_X(k)| \le \int_S P_X(x) \mathrm{d}x = 1$$

And we have that:

$$\varphi_X(k + \mathrm{d}k) = \int_S e^{i(k+\mathrm{d}k)x} P_X(x) \mathrm{d}x = \int_S e^{ix\mathrm{d}k} e^{ikx} P_X(x) \mathrm{d}x$$

Now since (as seen above) the integral converges we can switch the limit and the integral and we get:

$$\lim_{\mathrm{d}k \to 0} \varphi_X(k + \mathrm{d}k) = \int_S \lim_{\mathrm{d}k \to 0} e^{ix\mathrm{d}k} e^{ikx} P_X(x) \mathrm{d}x = \varphi_X(k)$$

Then notice that:

$$\varphi_X(k) = (\varphi_X(-k))^\star$$

2. We saw from the previous question that:

$$\varphi_X(k) = (\varphi_X(-k))^\star = \varphi_X(k)^\star \Rightarrow \varphi_X(k) \in \mathbb{R} \quad \text{and} \quad \varphi_X(k) = \varphi_X(-k)$$

3. We have that φ_X must be a real valued even function. Then furthermore we have that:

$$\varphi_{\alpha X + \beta Y} = \varphi_{(\alpha+\beta)X} \Leftrightarrow \varphi_X(\alpha \cdot k) + \varphi_X(\beta \cdot k) = \varphi_X((\alpha + \beta) \cdot k)$$

This is a case of Cauchy's functional equation which since φ_X must be Lebesgue measurable admits only linear functions as solutions over $\mathbb{R}$. Then we have that:

$$\varphi_X(k) = ck???$$

6.3 Convergence in law

1. We have that:
$$\forall f \text{ continous and bounded}, E[f(X_n)] \to E[f(X)]$$
Hence if f is continuous we can define:
$$\forall g \text{ continous and bounded}, E[(g \circ f)(X_n)] = E[(g \circ f)(X)]$$
Since $g \circ f$ is continuous and bounded.
2. We want:
$$\forall \varepsilon > 0, P(|X_n - c| > \varepsilon) = 0$$
However from the Markov inequality we have that:
$$P(|X_n - c| > \varepsilon) \leq \frac{1}{\varepsilon} E[|X_n - c|] = \frac{1}{\varepsilon} E[|X - c|] = 0$$
Where we used the convergence in distribution in the first equality.

6.4 Stirling's formula

1.
2.
3. From exercise 1 we know that $S_n \sim \text{Poisson}(n)$. Gaussian from law of large number.
4. ...
5. ...

Chapter 7

Extreme value distributions

1. We have that:

   $F_{M_n}(x) = P(M_n \leq x) = P(\max(X_1, \cdots, X_n) \leq x) = P(X_1 \leq x \wedge \cdots \wedge X_n \leq x) = P(X \leq x)^n = (F_X(x))^n$

2. (a)

   (b)

   (c)

3. We have that $F_{R_X}(x) = 1_{x \geq R_x} = 1_{F_X(x)=1}$ and from Question 1 we know that $F_{M_n}(x) = (F_X(x))^n \to 1_{F_X(x)=1}$. Hence $M_n \xrightarrow{d} R_X$.

4. (a) We have that:

   $$F_{\hat{M}_n}(\hat{x}) = F_{M_n}(b_n\hat{x} + a_n) = (F_X(b_n\hat{x} + a_n))^n$$

   (b) Using the previous question and assuming the form of the a_n and b_n we have that:

   $$F_{\hat{M}_n}(\hat{x}) = \left(1 - \frac{\gamma(\hat{x})}{n} + o\left(\frac{1}{n}\right)\right)^n \overset{n\to\infty}{\longrightarrow} e^{-\gamma(\hat{x})}$$

5. We want:

   $$F_X(a_n + \hat{x}b_n) = 1 - \frac{e^{-\hat{x}}}{n} + o(\frac{1}{n})$$

   Notice that:

   $$F_X(a_n + \hat{x}b_n) = (1 - e^{-a_n - \hat{x}b_n})\delta_{a_n + \hat{x}b_n \geq 0}$$

   Hence taking $b_n = 1$ and $a_n = \log n$ gives the desired result.

6. Again we get:

   $$F_X(a_n + \hat{x}b_n) = 1 - (1 - a_n - \hat{x}b_n)^\alpha \quad \text{when} \quad a_n + \hat{x}b_n \in [0, 1]$$

   Now taking $a_n = 1$ and $b_n = n^{-\frac{1}{\alpha}}$ we get indeed that:

   $$F_{M_n}(\hat{x}) \to e^{-(-\hat{x})^\alpha} \quad \text{when} \quad \hat{x} \leq 0$$

   And $a_n + \hat{x}b_n \geq 1 \Leftrightarrow \hat{x} \geq 0$ giving the second part.

7. Similarly:

   $$F_X(a_n + \hat{x}b_n) = 1 - (a_n + \hat{x}b_n)^{-\alpha}$$

   Then taking $a_n = 0$ and $b_n = n^{\frac{1}{\alpha}}$ we get the desired result.

8. Again we have:

   $$F_X(a_n + \hat{x}b_n) = 1 - \frac{1}{a_n + \hat{x}b_n}\frac{1}{\sqrt{2\pi}}e^{-\frac{(a_n + \hat{x}b_n)^2}{2}} + O\left(\frac{e^{-\frac{(a_n+\hat{x}b_n)^2}{2}}}{x^3}\right)$$

   Now looking at this term by term notice that:

   $$e^{-\frac{(a_n+\hat{x}b_n)^2}{2}} = e^{-\ln n} \cdot e^{\cdots} \cdot e^{-\frac{\hat{x}^2}{4\ln n}} \cdot e^{-\hat{x}(1 + \frac{\ln\ln n + \ln 4\pi}{4\ln n})}$$

   Finishing the computations will give the desired result.

9. Since $F_{\max(X_1,X_2)}(x) = F_X(x)^2$ then if X_1, X_2 reached R_X with a power-law behavior (resp. inverse power-law and faster than power-law) then so will $\max(X_1, X_2)$ and therefore $\max(X_1, X_2)$ is in the same domain of attraction as X_1 and X_2.