Midterm Exam - April 24, 2018, Limit thms. of probab.

Family name ______________________ **Given name** ______________________

Signature ______________________ **Neptun Code** ______________________

No calculators or electronic devices are allowed. One formula sheet with 15 formulas is allowed.

1. (7 marks) Given some $p \in (0,1)$, let $Y_{p,0}, Y_{p,1}, Y_{p,2}, \dots$ denote i.i.d. random variables with Bernoulli distribution:
$$\mathbb{P}(Y_{p,n} = 1) = p, \qquad \mathbb{P}(Y_{p,n} = 0) = 1 - p.$$
Let $X_p = \min\{ n \geq 0 \, : \, Y_{p,n} = 1 \}$.
Use the method of characteristic functions to prove that $pX_p \Rightarrow \text{EXP}(1)$ as $p \to 0_+$.
Solution: X_p has pessimistic geometric distribution: $\mathbb{P}(X_p = k) = (1-p)^k p, \; k = 0, 1, 2, \dots$
$\varphi_p(t) = \mathbb{E}(e^{itX_p}) = \sum_{k=0}^\infty e^{itk}(1-p)^k p = p \sum_{k=0}^\infty (e^{it}(1-p))^k = \frac{p}{1-e^{it}(1-p)}$.
$\lim_{p \to 0} \varphi_p(pt) = \lim_{p \to 0} \frac{p}{1-e^{ipt}(1-p)} = \lim_{p \to 0} \frac{1}{1-(e^{ipt}-1)/p} = \frac{1}{1-it}$, the characteristic function of EXP(1).

2. (8 marks) Let $1 > p_1 \geq p_2 \geq p_3 \geq \dots \geq 0$. Let $X_1, X_2, \dots$ denote independent random variables with Bernoulli distribution:
$$\mathbb{P}(X_n = 1) = p_n, \qquad \mathbb{P}(X_n = 0) = 1 - p_n.$$
Let us define $S_n = X_1 + \dots + X_n$. Write down the extra conditions that we need to impose on the sequence $(p_n)_{n=1}^\infty$ so that we can conclude that
$$\frac{S_n - \sum_{k=1}^n p_k}{\sqrt{\sum_{k=1}^n p_k}} \Rightarrow \mathcal{N}(0,1).$$
Hint: Use Lindeberg's theorem.
Solution:
$\mathbb{E}(S_n) = \sum_{k=1}^n p_k, \;\; \text{Var}(S_n) = \sum_{k=1}^n p_k(1-p_k)$.
In order to prove $\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \Rightarrow \mathcal{N}(0,1)$, we need to check Lindeberg's condition. We want to show that for any $\varepsilon > 0$ we have
$$\lim_{n \to \infty} \frac{1}{\text{Var}(S_n)} \sum_{k=1}^n \mathbb{E}\left((X_k - p_k)^2 \mathbb{1}\left[|X_k - p_k| > \varepsilon \sqrt{\text{Var}(S_n)} \right] \right) = 0$$
If $\lim_{n \to \infty} \text{Var}(S_n) = +\infty$ then $\mathbb{1}\left[|X_k - p_k| > \varepsilon\sqrt{\text{Var}(S_n)} \right] = 0$ for all $k \in \mathbb{N}$ and all $n \geq n_0$, where n_0 is the smallest index for which $\varepsilon\sqrt{\text{Var}(S_{n_0})} > 1$, since $|X_k - p_k| \leq 1$ for any k. Thus we have
$$\frac{1}{\text{Var}(S_n)} \sum_{k=1}^n \mathbb{E}\left((X_k - p_k)^2 \mathbb{1}\left[|X_k - p_k| > \varepsilon \sqrt{\text{Var}(S_n)} \right] \right) = 0, \qquad n \geq n_0.$$
Now we note that if $\sum_{k=1}^\infty p_k = +\infty$ then $\lim_{n \to \infty} \text{Var}(S_n) = \lim_{n \to \infty} \sum_{k=1}^n p_k(1-p_k) = +\infty$, thus we have $\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}} \Rightarrow \mathcal{N}(0,1)$ by Lindeberg's theorem. In order to conclude $\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\sum_{k=1}^n p_k}} \Rightarrow \mathcal{N}(0,1)$, we need
$$\lim_{n \to \infty} \frac{\sqrt{\sum_{k=1}^n p_k(1-p_k)}}{\sqrt{\sum_{k=1}^n p_k}} = 1.$$

The necessary and sufficient condition for this is $\lim_{n \to \infty} p_n = 0$ (in addition to $\sum_{k=1}^\infty p_k = +\infty$).

Also note that if $\sum_{k=1}^\infty p_k < +\infty$ then $\mathbb{E}(S_\infty) < +\infty$, thus $\mathbb{P}(S_\infty < +\infty) = 1$, thus in this case we actually have $\lim_{n \to \infty} \frac{S_n - \sum_{k=1}^n p_k}{\sqrt{\sum_{k=1}^n p_k}} = \frac{S_\infty - \sum_{k=1}^\infty p_k}{\sqrt{\sum_{k=1}^\infty p_k}}$, and the limiting random variable is discrete, so it definitely doesn't have standard normal distribution.