Department of Mathematics, IST — Section of Probability and Statistics

Probability Theory

2nd. TEST 1st. Semester — 2010/11

Duration: 1h30m 2011/01/31 — **5PM**, Room P2

- Please justify your answers.
- This test has two pages and four groups. The total of points: 20.0.

Group V — Independence and Bernoulli/Poisson processes **6.0 points**

1. *A Bernoulli process with parameter p has already been used in the investigation of earths magnetic field reversals,[1] with Bernoulli trials separated by 282 ky (i.e. 282 thousand years).* (2.5)

   Prove that, given that the 4th. geomagnetic reversal occurred at the 100th. Bernoulli trial (that is, $\{T_4 = 100\}$), the joint distribution of (T_1, T_2, T_3), the vector of the number of 282 ky periods until the 1st., 2nd. and 3rd. geomagnetic reversals, is the same as the distribution of an ordered random sample of 3 numbers chosen without replacement from $\{1, 2, \ldots, 99\}$.

   - **Bernoulli process**

     $\{X_i,\ i \in \mathbb{N}\} \stackrel{i.i.d.}{\sim} \text{Bernoulli}(p)$

     $$X_i = \begin{cases} 1, & \text{if there is a geomagnetic reversal during the } i^{th} \text{ 282 ky period} \\ 0, & \text{otherwise} \end{cases}$$

   - **R.v.**

     According to Prop. 3.83,

     $$\begin{aligned} T_k &= \text{number of 282 ky periods until the } k^{th} \text{geomagnetic reversal} \\ &\sim \text{NegativeBinomial}(k, p),\ k \in \mathbb{N}. \end{aligned}$$

   - **Requested probability**

     For $1 \le t_1 < t_2 < t_3 < t_4 = 100$,

     $$P(T_1 = t_1, T_2 = t_2, T_3 = t_3 \mid T_4 = 100) = \frac{P(T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = 100)}{P(T_4 = 100)},$$

     where

     $$\begin{aligned} P(T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = 100) &= P\Big[\big(\cap_{i=1}^{t_1-1}\{X_i = 0\}\big) \cap X_{t_1} = 1, \\ &\quad \big(\cap_{i=t_1+1}^{t_2-1}\{X_i = 0\}\big) \cap X_{t_2} = 1, \\ &\quad \big(\cap_{i=t_2+1}^{t_3-1}\{X_i = 0\}\big) \cap X_{t_3} = 1, \\ &\quad \big(\cap_{i=t_3+1}^{100-1}\{X_i = 0\}\big) \cap X_{100} = 1\Big] \end{aligned}$$

[1] *A geomagnetic reversal is a change in the orientation of Earth's magnetic field such that the positions of magnetic north and magnetic south become interchanged (http://en.wikipedia.org/wiki/Geomagnetic_reversal).*

$$\begin{aligned} P(T_1 = t_1, T_2 = t_2, T_3 = t_3, T_4 = 100) &= \left[(1-p)^{t_1-1}\, p\right] \times \left[(1-p)^{t_2-t_1-1}\, p\right] \\ &\quad \times \left[(1-p)^{t_3-t_2-1}\, p\right] \times \left[(1-p)^{100-t_3-1}\, p\right] \\ &= (1-p)^{100-4}\, p^4 \end{aligned}$$

and

$$P(T_4 = 100) = \binom{100-1}{4-1} p^4\, (1-p)^{100-4}.$$

Thus,

$$P(T_1 = t_1, T_2 = t_2, T_3 = t_3 \mid T_4 = 100) = \frac{1}{\binom{99}{3}},$$

which is indeed the distribution of a random sample of 3 numbers chosen without replacement from $\{1, 2, \ldots, 99\}$.

2. *Suppose that the number of requests to a web server follows a conditional (or mixed) Poisson process with random rate Λ (in requests per minute) and admit that $\Lambda \sim Gamma(r, \beta)$, where $r, \beta \in I\!N$.* (3.5)

   Derive a simplified expression for the probability that the server receives at most m requests in t minutes ($m \in I\!N_0$, $t > 0$) and obtain the value of this probability for $r = \beta = t = 2$, $m = 8$.

   - **Stochastic process**

     $\{N(t), t \ge 0\}$ conditional (or mixed) Poisson process with random rate Λ

   - **R.v.**

     $N(t)$ = number of requests to a web server in the 1st. t minutes

     $(N(t) \mid \Lambda = \lambda) \sim \text{Poisson}(\lambda t)$

   - **Random rate and its p.d.f.**

     $\Lambda \sim \text{Gamma}(r, \beta),\ r, \beta \in \mathbb{N}$

     $g_\Lambda(\lambda) = \frac{\beta^r}{\Gamma(r)}\, \lambda^{r-1}\, e^{-\beta\lambda},\ \lambda \ge 0$

   - **P.f. of $N(t)$**

     Since Λ is a continuous r.v. with p.d.f. $g_\Lambda(\lambda)$, we get

     $$P[N(t+s) - N(s) = n] = \int_0^{+\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!}\, g_\Lambda(\lambda) d\lambda.$$

     by Prop. 3.132. Moreover, since $\{N(t), t \ge 0\}$ has stationary increments, we can add that:

     $$\begin{aligned} P[N(t) = n] &= \int_0^{+\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \frac{\beta^r}{\Gamma(r)} \lambda^{r-1} e^{-\beta\lambda} d\lambda \\ &= \frac{t^n \Gamma(n+r)\beta^r}{n!\,\Gamma(r)(\beta+t)^{r+n}} \int_0^{+\infty} \frac{(\beta+t)^{n+r}}{\Gamma(n+r)} \lambda^{n+r-1} e^{-(\beta+t)\lambda}\, d\lambda \\ &= \frac{t^n \Gamma(n+r)\beta^r}{n!\,\Gamma(r)(\beta+t)^{r+n}} \int_0^{+\infty} f_{Gamma(n+r, \beta+t)}(\lambda)\, d\lambda \end{aligned}$$

$$
\begin{aligned}
P[N(t)=n] &= \binom{n+r-1}{n}\left(\frac{\beta}{\beta+t}\right)^r\left(1-\frac{\beta}{\beta+t}\right)^n,\ n\in\mathbb{N}_0\\
&\equiv \text{p.f. of a NegativeBinomial}^*(r,\beta/(\beta+t)).
\end{aligned}
$$

- **Requested probability**

  For $m \in \mathbb{N}_0$, we have

$$
\begin{aligned}
P[N(t)\le m] &= F_{NegativeBinomial^*(r,\beta/(\beta+t))}(m)\\
&= F_{NegativeBinomial(r,\beta/(\beta+t))}(m+r)\\
&= 1-F_{Binomial(m+r,\beta/(\beta+t))}(r-1).
\end{aligned}
$$

  Thus, for $r=\beta=t=2$ and $m=8$, we get

$$
\begin{aligned}
P[N(2)\le 8] &= 1-F_{Binomial(8+2,2/(2+2))}(2-1)\\
&\stackrel{table}{=} 1-0.0107\\
&= 0.9893.
\end{aligned}
$$

Group VI — (In)dependence and expectation **4.5 points**

1. *Let X and Y be L^2 r.v. with zero mean, unit variance and covariance ρ.* (2.5)

   Show that $E\left(\max\{X^2,Y^2\}\right)\le 1+\sqrt{1-\rho^2}$.

   Hint: *Use the identity* $\max\{u,v\}=\frac{1}{2}(u+v)+\frac{1}{2}|u-v|$ *and the Cauchy-Schwarz inequality.*

- **Random vector**

  (X,Y), where $X,Y\in L^2$ and

  (i) $E(X)=E(Y)=0$

  (ii) $V(X)=V(Y)=1$

  (iii) $cov(X,Y)=\rho$.

- **Auxiliary results**

$$
\text{(i) \& (ii)} \ \Rightarrow\ \begin{cases} E(X^2)=V(X)+E^2(X)=V(X)=1\\ E(Y^2)=V(Y)+E^2(Y)=V(Y)=1\end{cases}
$$

$$
\text{(i) \& (ii) \& (iii)} \ \Rightarrow\ cov(X,Y)=E(XY)-E(X)\,E(Y)=E(XY)=\rho
$$

- **To prove**

  $E\left(\max\{X^2,Y^2\}\right)\le 1+\sqrt{1-\rho^2}$

- **Proof**

  Using the hint ($\max\{u,v\}=\frac{1}{2}(u+v)+\frac{1}{2}|u-v|$), we get:

$$
\begin{aligned}
E\left(\max\{X^2,Y^2\}\right) &= E\left[\frac{1}{2}(X^2+Y^2)+\frac{1}{2}|X^2-Y^2|\}\right]\\
&= \frac{1}{2}\left[E(X^2)+E(Y^2)\right]+\frac{1}{2}E\left[|(X-Y)(X+Y)|\right].
\end{aligned}
$$

  Now, note that $X,Y\in L^2$ and so $X-Y,X+Y\in L^2$. Hence, we can apply the Cauchy-Schwarz inequality to get:

$$
\begin{aligned}
E\left(\max\{X^2,Y^2\}\right) &= \frac{1}{2}(1+1)+\frac{1}{2}E\left[|(X-Y)(X+Y)|\right]\\
&\le 1+\frac{1}{2}\sqrt{E\left[(X-Y)^2\right]\,E\left[(X+Y)^2\right]}\\
&= 1+\frac{1}{2}\sqrt{\left[E(X^2)-2E(XY)+E(Y^2)\right]\left[E(X^2)+2E(XY)+E(Y^2)\right]}\\
&= 1+\frac{1}{2}\sqrt{(1-2\rho+1)(1+2\rho+1)}\\
&= 1+\frac{1}{2}\sqrt{4(1-\rho)(1+\rho)}\\
&= 1+\sqrt{1-\rho^2}.
\end{aligned}
$$

  QED

2. *The lifetimes of 2 computer systems are assumed to be i.i.d. r.v. with Weibull($\alpha=1,\beta=0.5$) distribution, where α and β are the scale and form parameters (respectively).* (2.0)

   What is the variance of the lifetime of the system that survives the shortest?

   Hint: *Capitalize on the closure of the Weibull distribution under the minimum operation.*

- **R.v.**

  $X_i=$ time to failure of component i, $i=1,2$

  $X_i \stackrel{i.i.d.}{\sim} X$, $i=1,2$

  $X\sim$ Weibull($\alpha=1$, $\beta=0.5$)

- **New r.v.**

  $X_{(1)}=\min\{X_1,X_2\}=$ lifetime of the system that survives the shortest

- **Survival function of $X_{(1)}$**

  Capitalizing once again on Example 3.65, we get

$$
\begin{aligned}
S_{X_{(1)}}(x) &= P[X_{(1)}>x]\\
&= [S_X(x)]^2\\
&\stackrel{form}{=} \left[e^{-\left(\frac{x}{\alpha}\right)^\beta}\right]^2\\
&= \exp\left[-\left(\frac{x}{\frac{\alpha}{2^{1/\beta}}}\right)^\beta\right]\\
&\equiv S_{Weibull(\alpha^*,\beta^*)}(x),
\end{aligned}
$$

  where, for $\alpha=1$ and $\beta=0.5$,

$$
\begin{aligned}
\alpha^* &= \frac{\alpha}{2^{1/\beta}}\\
&= \frac{1}{4}\\
\beta^* &= \beta\\
&= 0.5.
\end{aligned}
$$

- **Variance of $X_{(1)}$**

$$\begin{aligned} V[X_{(1)}] &\overset{form}{=} (\alpha^*)^2 \left[\Gamma\left(1+\frac{2}{\beta^*}\right) - \Gamma^2\left(1+\frac{1}{\beta^*}\right)\right] \\ &= \left(\frac{1}{4}\right)^2 \left[\Gamma(5) - \Gamma^2(3)\right] \\ &= \frac{1}{16}\left[4! - (2!)^2\right] \\ &= \frac{20}{16} \\ &= 1.25. \end{aligned}$$

Group VII — Expectation **5.5 points**

1. *Suppose that a customer arrives at a store, which sells a single type of commodity. The amount desired by this customer is assumed to be a r.v. X with d.f.* (3.0)

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1-e^{-\lambda x}}{1-e^{-\lambda S}}, & 0 \le x \le S \\ 1, & x > S, \end{cases}$$

where S is a positive constant that represents the maximum amount desired by this customer.

The store uses the following (s, S) ordering policy $(0 < s < S)$:

- *if the inventory level after serving a customer is below s, then an order is placed to bring it up to level S;*
- *otherwise no order is placed.*

Find a simplified expression for the expected value of the r.v. Y that represents the amount ordered after serving this customer.

Hint: *Derive the d.f. of Y.*

- **R.v.**

  $X =$ amount desired by the customer

- **D.f. of X**

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1-e^{-\lambda x}}{1-e^{-\lambda S}}, & 0 \le x \le S \\ 1, & x > S, \end{cases}$$

- **Range of X**

  $\mathbb{R}_X = [0, S]$

- **New r.v.**

  $Y =$ amount ordered after serving the customer

- **Relating Y and X**

  According to the (s, S) ordering policy $(0 < s < S)$:

  - if the inventory level after serving a customer is below s, then an order is placed to bring it up to level S;
  - otherwise no order is placed.

  Hence,

$$Y = \begin{cases} 0, & 0 \le X \le S-s \\ X, & S-s < X \le S, \end{cases}$$

- **Range of Y**

  $\mathbb{R}_Y = \{0\} \cup (S-s, S]$

- **D.f. of Y**

  The d.f. of Y, $F_Y(y)$, is defined as follows:

  - for $y < 0$,

$$P(Y \le y) = 0;$$

  - for $y = 0$,

$$\begin{aligned} P(Y \le 0) &= P(0 \le X \le S-s) \\ &= F_X(S-s) \\ &= \frac{1-e^{-\lambda(S-s)}}{1-e^{-\lambda S}}; \end{aligned}$$

  - for $0 < y \le S-s$,

$$\begin{aligned} P(Y \le y) &= P(Y \le 0) \\ &= \frac{1-e^{-\lambda(S-s)}}{1-e^{-\lambda S}}; \end{aligned}$$

  - for $S-s < y \le S$,

$$\begin{aligned} P(Y \le y) &= P(Y = 0) + P(S-s < X \le y) \\ &= P(0 < X \le y) \\ &= \frac{1-e^{-\lambda y}}{1-e^{-\lambda S}}; \end{aligned}$$

  - for $y > S$,

$$P(Y \le y) = 1.$$

- **Simplified expression of $E(Y)$**

  $Y \ge 0$, thus, by Th. 4.65,

$$\begin{aligned} E(Y) &= \int_0^{+\infty} [1 - F_Y(y)]\, dy \\ &= \int_0^{S-s} \left[1 - \frac{1-e^{-\lambda(S-s)}}{1-e^{-\lambda S}}\right] dy + \int_{S-s}^{S} \left[1 - \frac{1-e^{-\lambda y}}{1-e^{-\lambda S}}\right] dy \\ &= \int_0^{S-s} \frac{e^{-\lambda(S-s)} - e^{-\lambda S}}{1-e^{-\lambda S}}\, dy + \int_{S-s}^{S} \frac{e^{-\lambda y} - e^{-\lambda S}}{1-e^{-\lambda S}}\, dy \end{aligned}$$

$$
\begin{aligned}
E(Y) &= \int_0^{S-s} \frac{e^{-\lambda(S-s)}}{1-e^{-\lambda S}}\,dy + \int_{S-s}^{S} \frac{e^{-\lambda y}}{1-e^{-\lambda S}}\,dy + \int_0^{S} \frac{e^{-\lambda S}}{1-e^{-\lambda S}}\,dy \\
&= (S-s) \times \frac{e^{-\lambda(S-s)}}{1-e^{-\lambda S}} - \frac{1}{\lambda}\left(\left.\frac{e^{-\lambda y}}{1-e^{-\lambda S}}\right|_{S-s}^{S} \right) + S \times \frac{e^{-\lambda S}}{1-e^{-\lambda S}} \\
&= (S-s) \times \frac{e^{-\lambda(S-s)}}{1-e^{-\lambda S}} + \frac{e^{-\lambda(S-s)} - e^{-\lambda S}}{\lambda\,[1-e^{-\lambda S}]} + S \times \frac{e^{-\lambda S}}{1-e^{-\lambda S}} \\
&\overset{Mathematica}{=} \frac{e^{\lambda s}(\lambda s - \lambda S - 1) + \lambda S e^{\lambda S} - \lambda S + 1}{\lambda\,(e^{\lambda S}-1)}.
\end{aligned}
$$

2. *Suppose that X_1 and X_2 are independent normal random variables with the same variance.* (2.5)
Show that $X_1 - X_2$ and $X_1 + X_2$ are independent r.v.

- **R.v.**
$X_i \overset{indep.}{\sim} \text{Normal}(\mu_1, \sigma_i^2),\ i = 1, 2$

- **Joint distribution**
Consider

$$
\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \qquad
\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.
$$

Then

$$
(X_1, X_2) \sim \text{Normal}_2(\underline{\mu}, \mathbf{\Sigma}),
$$

according to Def. 4.181.

- **Transformation**

$$
\begin{bmatrix} X_1 - X_2 \\ X_1 + X_2 \end{bmatrix} = \mathbf{C}\underline{X} + \underline{b}
$$

where

$$
\mathbf{C} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \qquad
\underline{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.
$$

- **Distribution of $X_1 + X_2$**
According to Th. 4.190,

$$
X_1 + X_2 \sim \text{Normal}_2(\mathbf{C}\,\underline{\mu} + \underline{b}, \mathbf{C}\,\mathbf{\Sigma}\,\mathbf{C}^\top),
$$

where:

$$
\begin{aligned}
\mathbf{C}\,\underline{\mu} + \underline{b} &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 + \mu_2 \end{bmatrix};
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{C}\,\mathbf{\Sigma}\,\mathbf{C}^\top &= \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \times \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \times \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \sigma_1^2 + \sigma_2^2 & \sigma_1^2 + \sigma_2^2 \\ \sigma_1^2 - \sigma_2^2 & \sigma_1^2 + \sigma_2^2 \end{bmatrix} \times \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\
&\overset{\sigma_1=\sigma_2=\sigma}{=} \begin{bmatrix} 2\sigma^2 & 0 \\ 0 & 2\sigma^2 \end{bmatrix}.
\end{aligned}
$$

- **Verifying the independence between $X_1 - X_2$ and $X_1 + X_2$**
We have just conclude that if $\sigma_1 = \sigma_2 = \sigma$ then $(X_1 - X_2, X_1 + X_2)$ has a bivariate normal distribution with $cov(X_1 - X_2, X_1 + X_2) = 0$. Thus,

$$
X_1 - X_2 \perp\!\!\!\perp X_1 + X_2,
$$

according to Cor. 4.192. QED

Group VIII — Convergence of sequences of r.v. **4.0 points**

1. *Let $X_1, X_2, \ldots$ be a sequence of r.v. in L^2 and X a r.v. also in L^2.* (2.0)
Prove that $X_n \overset{q.m.}{\to} X \Rightarrow E(X_n^2) \to E(X^2)$.
Hint: *Use the fact that $X_n^2 = (X_n - X)^2 + 2X(X_n - X) + X^2$ and the modulus and Cauchy-Schwarz inequalities.*

- **Sequence of r.v.**
$\{X_1, X_2, \ldots\}$
$X_i \in L^2$

- **To prove**
$X_n \overset{q.m.}{\to} X \Rightarrow E(X_n^2) \to E(X^2)$, where $X \in L^2$

- **Proof**
Let us remind the reader that, according to Def. 5.15,

$$
X_n \overset{q.m.}{\to} X \Leftrightarrow \lim_{n\to+\infty} E[(X_n - X)^2] = 0. \tag{1}
$$

Since

$$
X_n^2 = (X_n - X)^2 + 2X(X_n - X) + X^2,
$$

where $X_n - X \in L^2$, we get:

$$\begin{aligned}
\lim_{n\to+\infty} E(X_n^2) &= \lim_{n\to+\infty} E[(X_n - X)^2 + 2X(X_n - X) + X^2] \\
&= \lim_{n\to+\infty} \left\{ E[(X_n - X)^2] + 2\,E[X(X_n - X)] + E(X^2) \right\} \\
&\overset{(1)}{=} 2 \lim_{n\to+\infty} E[X(X_n - X)] + E(X^2).
\end{aligned}$$

Moroever, the Cauchy-Schwarz inequality leads to

$$\begin{aligned}
0 \le \lim_{n\to+\infty} E[|X(X_n - X)|] &\le \lim_{n\to+\infty} \sqrt{E(X^2)\,E[(X_n - X)^2]} \\
&= E(X^2) \sqrt{\lim_{n\to+\infty} E[(X_n - X)^2]} \\
&= 0,
\end{aligned}$$

i.e.,

$$\lim_{n\to+\infty} E[|X(X_n - X)|] = 0. \tag{2}$$

Furthermore, by the modulus inequality (Cor. 4.43),

$$0 = \lim_{n\to+\infty} E[|X(X_n - X)|] \ge \lim_{n\to+\infty} |E[X(X_n - X)]|, \tag{3}$$

hence,

$$\lim_{n\to+\infty} E[X(X_n - X)] = 0.$$

We can finally state that

$$\begin{aligned}
\lim_{n\to+\infty} E(X_n^2) &= 2 \times 0 + E(X^2) \\
&= E(X^2).
\end{aligned}$$

QED

2. *Consider $\{X_1, X_2, \dots\}$ a sequence of independent Poisson r.v. with respective parameters $\{\lambda_1, \lambda_2, \dots\}$.* (2.0)

Show that if $\sigma^2 = \sum_{i=1}^{+\infty} \lambda_i < +\infty$ then $S_n = \sum_{i=1}^{n} X_i \overset{a.s.}{\to} S = \sum_{i=1}^{+\infty} X_i$, where $S \sim$ Poisson(σ^2).

Hint: *Proving that $\sup_{k\ge n} |S_n - S| \overset{P}{\to} 0$ is quite useful.*

- **Sequence of r.v.**

  $\{X_1, X_2, \dots\}$

  $X_i \overset{indep.}{\sim}$ Poisson(λ_i), $i = 1, 2, \dots$

- **Another sequence of r.v.**

  $\{S_1, S_2, \dots\}$

  $S_n = \sum_{i=1}^{n} X_i \overset{Prop. 3.55}{\sim}$ Poisson $\left(\sum_{i=1}^{n} \lambda_i\right)$, $i = 1, 2, \dots$

- **To prove**

  $\sum_{i=1}^{+\infty} \lambda_i = \sigma^2 < +\infty \;\Rightarrow\; S_n = \sum_{i=1}^{n} X_i \overset{a.s.}{\to} S = \sum_{i=1}^{+\infty} X_i \sim$ Poisson(σ^2)

- **Proof**

  According to Prop. 5-37 (relationship between almost sure convergence and convergence in probability),

$$S_n \overset{a.s.}{\to} S \quad\Leftrightarrow\quad Y_n = \sup_{k\ge n} |S_k - S| \overset{P}{\to} 0,$$

that is, we only need to prove that, for any $\epsilon > 0$,

$$\lim_{n\to+\infty} P(|Y_n - 0| > \epsilon) = 0.$$

Firstly, note that

$$\begin{aligned}
Y_n &= \sup_{k\ge n} |S_k - S| \\
&\overset{X_i\ge 0}{=} \sup_{k\ge n} \sum_{i=k+1}^{+\infty} X_i \\
&\overset{X_i\ge 0}{=} \sum_{i=n+1}^{+\infty} X_i \\
&\overset{\sigma^2=\sum_{i=1}^{+\infty}\lambda_i<+\infty}{\sim} \text{Poisson}\left(\sum_{i=n}^{+\infty} \lambda_i\right) \\
E(Y_n) &= V(Y_n) \\
&= \sum_{i=n}^{+\infty} \lambda_i.
\end{aligned}$$

Secondly, we ought to mention that since $\sum_{i=1}^{+\infty} \lambda_i = \sigma^2 < +\infty$,

$$\lim_{n\to+\infty} E(Y_n) = \lim_{n\to+\infty} V(Y_n) = 0.$$

Thirdly, if we apply Chebyshev's inequality (for non-negative r.v. and $g(x) = x^2$) and capitalise on the previous result then

$$\begin{aligned}
\lim_{n\to+\infty} P(Y_n > \epsilon) &\le \lim_{n\to+\infty} \frac{E(Y_n^2)}{\epsilon^2} \\
&= \lim_{n\to+\infty} \frac{V(Y_n) + E^2(Y_n)}{\epsilon^2} \\
&= \frac{0+0}{\epsilon^2} \\
&= 0,
\end{aligned}$$

that is, $Y_n = \sup_{k\ge n} |S_k - S| \overset{P}{\to} 0$ and, thus, $S_n \overset{a.s.}{\to} S$. QED