Notes 9 : CLT and Poisson Convergence

Math 733-734: Theory of Probability
Lecturer: Sebastien Roch

References: [Dur10, Section 3.4, 3.6].

1 Deterministic Lemmas

We will need some deterministic lemmas throughout.

LEM 9.1 Let \( z_1, \ldots, z_n \) and \( w_1, \ldots, w_n \) be complex numbers of modulus \( \leq \theta \). Then

\[
\left| \prod_{m=1}^{n} z_m - \prod_{m=1}^{n} w_m \right| \leq \theta^{n-1} \sum_{m=1}^{n} |z_m - w_m|.
\]

Proof:

\[
\left| \prod_{m=1}^{n} z_m - \prod_{m=1}^{n} w_m \right| \leq \left| z_1 \prod_{m=2}^{n} z_m - \prod_{m=2}^{n} w_m \right| + \left| z_1 \prod_{m=2}^{n} w_m - w_1 \prod_{m=2}^{n} w_m \right|
\]

\[
\leq \theta \left| \prod_{m=2}^{n} z_m - \prod_{m=2}^{n} w_m \right| + \theta^{n-1} |z_1 - w_1|,
\]

and use induction.

LEM 9.2 If \( \max_{1 \leq j \leq n} |c_{j,n}| \to 0 \), \( \sum_{j=1}^{n} c_{j,n} \to \lambda \) and \( \sup_{n} \sum_{j=1}^{n} |c_{j,n}| < \infty \) then

\[
\prod_{j=1}^{n} (1 + c_{j,n}) \to e^{\lambda}.
\]

Proof: Note that \( \frac{\log(1+x)}{x} \to 1 \) as \( x \to 0 \). Hence \( \forall \varepsilon > 0, \exists \delta > 0 \) such that \( |x| < \delta \) implies

\[
x - \varepsilon |x| < \log(1 + x) < x + \varepsilon |x|.
\]

The following standard expansion is proved in [D].
LEM 9.3 We have
\[ \left| e^{ix} - \sum_{m=0}^{n} \frac{(ix)^m}{m!} \right| \leq \min \left( \frac{|x|^{n+1}}{(n+1)!}, \frac{2|x|^n}{n!} \right). \]

LEM 9.4 If \( z \) is a complex number then
\[ |e^z - (1 + z)| \leq |z|^2 e^{|z|}. \]

Proof: By a Taylor expansion,
\[ |e^z - (1 + z)| \leq |z|^2/2! + |z|^3/3! + \cdots \]
\[ \leq |z|^2(1/2! + |z|/3! + \cdots) \]
\[ \leq |z|^2 e^{|z|}. \]

\[ \square \]

2 Easy laws

2.1 CLT

As we saw before, the behavior of \( \phi \) around 0 contains information about the tail/moments of \( \mu \):

THM 9.5 We have
\[ \left| \mathbb{E}\left[ e^{itX} \right] - \sum_{m=0}^{n} \frac{\mathbb{E}[(itX)^m]}{m!} \right| \leq \mathbb{E} \left[ \min \left\{ \frac{|tX|^{n+1}}{(n+1)!}, \frac{2|tX|^n}{n!} \right\} \right]. \]

Proof: This follows from Lemma 9.3. \( \square \)

We can now prove the CLT.

THM 9.6 Let \( (X_n) \) be IID with \( \mathbb{E}[X_1] = \mu \) and \( \text{Var}[X_1] = \sigma^2 < +\infty \). Then if \( S_n = \sum_{k \leq n} X_k \)
\[ Z_n = \frac{S_n - n\mu}{\sigma \sqrt{n}} \Rightarrow Z, \]
where \( Z \sim N(0, 1) \).

Proof: Suffices to prove the result for \( \mu = 0 \). Note that
\[ \phi_{X_1}(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2), \]
where the error term is \( \leq t^2 \mathbb{E}[|t||X|^3 \wedge 2|X|^2] \). The expression inside the expectation is dominated by \( 2X^2 \) which is integrable. So (DOM) implies that the expectation in the error term goes to 0 as \( t \to 0 \).

By independence

\[
\phi_{Z_n}(t) = \left( 1 - \frac{t^2}{2n} + o(t^2) \right)^n \to e^{-t^2/2}.
\]

The inversion formula and continuity theorem conclude the proof. (In fact, one must prove the above limit for complex numbers. This follows from Lemmas 9.1 and 9.4.)

### 2.2 Poisson convergence

**THM 9.7** Let \( X_n \) be binomial with parameters \( n \) and \( \lambda/n \), for \( \lambda > 0 \). Then \( X_n \Rightarrow Z \) where \( Z \) is Poisson with parameter \( \lambda \).

**Proof:** The CF of \( X_n \) is

\[
\phi_{X_n}(t) = \left( \frac{\lambda}{n} e^{it} + \left( 1 - \frac{\lambda}{n} \right) \right)^n \to \exp \left( \lambda(e^{it} - 1) \right),
\]

for all \( t \) as \( n \to +\infty \), by Lemmas 9.1 and 9.4.

### 3 Lindeberg-Feller CLT

**THM 9.8 (Lindeberg-Feller CLT)** For each \( n \), let \( X_{n,m}, 1 \leq m \leq n \), be independent with \( \mathbb{E}[X_{n,m}] = 0 \). Suppose

1. \( \sum_{m=1}^{n} \mathbb{E}[X_{n,m}^2] \to 1 \).

2. \( \forall \varepsilon > 0, \lim_n \sum_{m=1}^{n} \mathbb{E}[|X_{n,m}|^2; |X_{n,m}| > \varepsilon] = 0 \).

Then

\[
Z_n = \sum_{m=1}^{n} X_{n,m} \Rightarrow Z,
\]

as \( n \to \infty \) where \( Z \sim N(0,1) \).

In other words, a sum of a large number of small independent effects is approximately normal.
EX 9.9 To recover our previous CLT, take \( X_{n,m} = \frac{X_m}{\sqrt{n}} \). The first condition
is clearly satisfied. If \( \varepsilon > 0 \)
\[
\sum_{m=1}^{n} \mathbb{E}[|X_{n,m}|^2; |X_{n,m}| > \varepsilon] = n \mathbb{E}[|X_1/\sqrt{n}|^2; |X_1/\sqrt{n}| > \varepsilon]
= \mathbb{E}[|X_1|^2; |X_1| > \varepsilon\sqrt{n}] \to 0,
\]
by (DOM) and \( \mathbb{E}[X_1^2] < +\infty \).

Proof: Letting \( \phi_{n,m} \) be the CF of \( X_{n,m} \) and \( \sigma_{n,m}^2 = \mathbb{E}[X_{n,m}^2] \). It suffices to prove
\[
\prod_{m=1}^{n} \phi_{n,m}(t) \to e^{-t^2/2}.
\]
We will show this by proving two claims.

CLAIM 9.10
\[
\left| \prod_{m=1}^{n} (1 - t^2 \sigma_{n,m}^2/2) - e^{-t^2/2} \right| \to 0.
\]

CLAIM 9.11
\[
\left| \prod_{m=1}^{n} \phi_{n,m}(t) - \prod_{m=1}^{n} (1 - t^2 \sigma_{n,m}^2/2) \right| \to 0.
\]

1. Claim 9.10. Note that
\[
\sigma_{n,m}^2 \leq \varepsilon^2 + \mathbb{E}[|X_{n,m}|^2; |X_{n,m}| > \varepsilon],
\]
so by the second condition we have \( \max_{1 \leq m \leq n} \sigma_{n,m}^2 \to 0 \) (where the maximum over the second term is bounded by its sum). By the first condition,
\[
\sum_{m=1}^{n} -t^2 \sigma_{n,m}^2/2 \to -t^2/2.
\]
The result follows from Lemma 9.2 (or Lemmas 9.1 and 9.4).

2. Claim 9.11.
By Lemma 9.3 above (this calculation explains why we need the more sophisticated error term; o.w. the \( \varepsilon \) would not come out),
\[
|\phi_{n,m}(t) - (1 - t^2 \sigma_{n,m}^2/2)|
\leq \mathbb{E}[|tX_{n,m}|^3 \wedge 2|tX_{n,m}|^2]
\leq \mathbb{E}[|tX_{n,m}|^3; |X_{n,m}| \leq \varepsilon] + \mathbb{E}[2|tX_{n,m}|^2; |X_{n,m}| > \varepsilon]
\leq \varepsilon t^3 \mathbb{E}[|X_{n,m}|^2; |X_{n,m}| \leq \varepsilon] + 2t^2 \mathbb{E}[|X_{n,m}|^2; |X_{n,m}| > \varepsilon].
Note that both terms on the LHS are bounded by 1 in absolute value (for \( n \) large enough by the max bound above). (Note this is not uniform in \( t \), but for any fixed \( t \) one can choose \( n \) large enough so that the norm is less than 1.) So the sum over \( m \) converges to 0 and the claim follows from Lemma 9.1.

## 3.1 Examples

A good example of a triangular array is the following, which we studied as an application of Chebyshev’s inequality.

**EX 9.12 (Random permutations)** Any permutation can be decomposed into cycles. E.g., if \( \pi = [3, 9, 6, 8, 2, 1, 5, 4, 7] \), then \( \pi = (136)(2975)(48) \). In fact, a uniform permutation can be generated by following a cycle until it closes and starting from the smallest unassigned element, and so on. Let \( X_{n,k} \) be the indicator that the \( k \)-th element in this construction precedes the closure of a cycle. E.g., we have \( X_{9,3} = X_{9,7} = X_{9,9} = 1 \). The construction above implies that the \( X_{n,k} \)'s are independent and

\[
P[X_{n,j} = 1] = \frac{1}{n - j + 1}.
\]

That is because only one of the remaining elements closes the cycle. (To prove independence formally, show by induction on \( j \) that

\[
P[X_{n,i} = x_{n,i}, \forall i \leq j] = \prod_{i=1}^{j} P[X_{n,i} = x_{n,i}].
\]

Letting \( S_n \) be the number of cycles in \( \pi \) we have

\[
\mathbb{E}[S_n] = \sum_{j=1}^{n} \frac{1}{n - j + 1} \sim \log n,
\]

and

\[
\text{Var}[S_n] = \sum_{j=1}^{n} \text{Var}[X_{n,j}] = \sum_{j=1}^{n} \left( \frac{1}{n - j + 1} - \frac{1}{(n - j + 1)^2} \right) \sim \log n.
\]

Then we have

\[
\frac{S_n}{\log n} \to p 1 \quad \text{in fact} \quad \frac{S_n - \log n}{(\log n)^{1/2+\varepsilon}} \to p 0,
\]

by Chebyshev’s inequality.
On the other hand, defining
\[ Z_{n,j} = \frac{X_{n,j} - (n - j + 1)^{-1}}{\sqrt{\log n}}, \]
we get \( \mathbb{E}[Z_{n,j}] = 0 \), \( \sum_{j=1}^{n} \mathbb{E}[Z_{n,j}^2] \to 1 \), and for \( \varepsilon > 0 \)
\[ \sum_{j=1}^{n} \mathbb{E}[|Z_{n,j}|^2; |Z_{n,j}| > \varepsilon] \to 0, \]
since the sum is 0 as soon as \( (\log n)^{-1/2} < \varepsilon \). (Note that \( (n - j + 1)^{-1} \leq 1 \).) Hence,
\[ \frac{S_n - \log n}{\sqrt{\log n}} \Rightarrow Z, \]
where \( Z \sim N(0,1) \).

4 Law of rare events

4.1 First proof

THM 9.13 (Law of rare events) For each \( n \), let \( X_{n,m}, 1 \leq m \leq n \), be independent with \( \mathbb{P}[X_{n,m} = 1] = p_{n,m} \) and \( \mathbb{P}[X_{n,m} = 0] = 1 - p_{n,m} \) and \( \mathbb{P}[X_{n,m} \geq 2] = \varepsilon_{n,m} \). Suppose
1. \( \sum_{m=1}^{n} p_{n,m} \to \lambda > 0 \).
2. \( \max_{1 \leq m \leq n} p_{n,m} \to 0 \).
3. \( \sum_{m=1}^{n} \varepsilon_{n,m} \to 0 \).

Then
\[ S_n = \sum_{m=1}^{n} X_{n,m} \Rightarrow Z, \]
as \( n \to \infty \) where \( Z \sim \text{Poi}(\lambda) \).

Proof: Under the last assumption, the probability that any of the \( X_{n,m} \)'s is \( \geq 2 \) goes to 0 as \( n \to +\infty \). Hence, by the converging together lemma (proved in homework), it suffices to consider the case \( \varepsilon_{n,m} = 0 \).

1. We first compute the moment-generating function of the Poisson distribution. Note that
\[ \phi_Z(t) = \mathbb{E}[e^{itZ}] = \sum_{k \geq 0} \frac{e^{-\lambda} \lambda^k}{k!} e^{itk} = e^{-\lambda} e^{e^{it} \lambda} = \exp(\lambda(e^{it} - 1)). \]
2. We compute the moment-generating function of a Bernoulli. Note that
\[ \phi_X(n,m)(t) = \mathbb{E}[e^{itX_{n,m}}] = (1 - p_{n,m}) + p_{n,m}e^{it} = 1 + p_{n,m}(e^{it} - 1). \]

3. Since \( \sum_{m=1}^{n} p_{n,m} \rightarrow \lambda \), it suffices to prove
\[ \left| \exp \left( \sum_{m=1}^{n} p_{n,m}(e^{it} - 1) \right) - \prod_{m=1}^{n} [1 + p_{n,m}(e^{it} - 1)] \right| \rightarrow 0. \]

Note that
\[ |\exp(p(e^{it} - 1))| = \exp(\text{Re}(e^{it} - 1)) = \exp(\cos t - 1) \leq 1 \]
and
\[ |1 + p(e^{it} - 1)| = |(1 - p) + pe^{it}| \leq 1, \]
for \( p \in [0, 1] \). So from Lemmas 9.1 and 9.4 above, using that \( \max_{1 \leq m \leq n} p_{n,m} \leq 1/2 \) and \( |e^{it} - 1| \leq 2 \),
\[ \left| \exp \left( \sum_{m=1}^{n} p_{n,m}(e^{it} - 1) \right) - \prod_{m=1}^{n} [1 + p_{n,m}(e^{it} - 1)] \right| \leq \sum_{m=1}^{n} |\exp(p_{n,m}(e^{it} - 1)) - [1 + p_{n,m}(e^{it} - 1)]| \]
\[ \leq \sum_{m=1}^{n} p_{n,m}^{2} |e^{it} - 1|^{2} \]
\[ \leq 4 \left( \max_{1 \leq m \leq n} p_{n,m} \right) \sum_{m=1}^{n} p_{n,m} \]
\[ \rightarrow 0. \]

**EX 9.14** A typical application of the law of rare events is to approximate a binomial. Assume you have 365 students in class. The probability that none of them has their birthday today is roughly \( e^{-1} \).

### 4.2 Rate of convergence

Recall the following.
THM 9.15 The following are equivalent:
1. \( F_{X_n}(x) \to F_X(x) \) for all points of continuity of \( F_X \).
2. \( \mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)] \) for all \( f \in C_b(\mathbb{R}) \).
3. \( \mathbb{E}[e^{itX_n}] \to \mathbb{E}[e^{itX}] \) for all \( t \in \mathbb{R} \).

There are several ways of measuring how fast weak convergence occurs. For two PMs \( \mu, \nu \), the following definition gives a natural notion of distance
\[
\|\mu - \nu\|_D = \sup_{f \in \mathcal{D}} \left| \int f(x)\mu(dx) - \int f(x)\nu(dx) \right|
\]
where \( \mathcal{D} \) is a class of functions. The choice \( \mathcal{D} = \{ f : f = 1_{(-\infty, x]}, x \in \mathbb{R} \} \) leads to the Kolmogorov-Smirnov distance.

For the record, the following is a standard result refining the CLT. The proof is in [D].

THM 9.16 (Berry-Esseen theorem) Let \( (X_n)_n \) be IID with \( \mathbb{E}[X_1] = 0 \), \( \mathbb{E}[X_1^2] = \sigma^2 \), and \( \mathbb{E}[X_1^3] = \rho < \infty \). If \( F_n \) is the DF of \( (X_1 + \cdots + X_n)/\sigma \sqrt{n} \) and \( F \) is the DF of the standard normal, then
\[
\sup_x |F_n(x) - F(x)| \leq \frac{3\rho}{\sigma^3 \sqrt{n}}.
\]

For the Poisson convergence theorem, we will use a stronger notion of distance.

DEF 9.17 (Total variation distance) Let \( \mu, \nu \) be probability measure on \( (\Omega, \mathcal{F}) \). The total variation distance between \( \mu \) and \( \nu \) is defined as
\[
\|\mu - \nu\|_{TV} = \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|.
\]
Note this corresponds to taking \( \mathcal{D} = \{ f : f = 1_A, A \in \mathcal{F} \} \).

In the countable case, we give an equivalent definition.

LEM 9.18 Assume \( \Omega = S \) is countable and \( \mathcal{F} = 2^S \). Then
\[
\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)|.
\]

Proof: By the triangle inequality, for any \( A \subseteq \Omega \)
\[
\sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)| \geq |\mu(A) - \nu(A)| + |\mu(A^c) - \nu(A^c)| = 2|\mu(A) - \nu(A)|,
\]
with equality when \( A = \{ \omega : \mu(\omega) \geq \nu(\omega) \} \).
4.2.1 Poisson convergence by the coupling method

We prove the following refinement of the Poisson convergence theorem.

**THM 9.19** For some $n$, let $X_{n,m}$, $1 \leq m \leq n$, be independent with $\mathbb{P}[X_{n,m} = 1] = p_{n,m}$ and $\mathbb{P}[X_{n,m} = 0] = 1 - p_{n,m}$. Then

$$\|\mu_{S_n} - \mu_Z\|_{TV} \leq \sum_{m=1}^{n} p_{n,m}^2,$$

where $S_n = \sum_{m=1}^{n} X_{n,m}$ and $Z \sim \text{Poi}(\lambda = \sum_{m \leq n} p_{n,m})$.

We will use coupling to prove the previous theorem. We restrict ourselves to a countable space $\Omega = S$ and $\mathcal{F} = 2^S$. We let $\Delta(S)$ be the set of all PMs on $S$.

**DEF 9.20 (Coupling of RVs)** A coupling of $\mu, \nu \in \Delta(S)$ is a pair of $S$-valued RVs $(X, Y) \in S^2$ (defined on a joint probability space) such that $X \sim \mu$ and $Y \sim \nu$.

**EX 9.21** Let $S = \{0, 1\}$. Assume $\mu = \nu$. Then $X \sim \nu$, $Y \sim \nu$ independent defines a coupling. So does $X = Y$. If $\mu \neq \nu$, the latter is not possible. In order to maximize the probability that $\mathbb{P}[X = Y]$ one can choose $\mathbb{P}[X = Y = \omega] = \mu(\omega) \land \nu(\omega)$, $\mathbb{P}[X = 1, Y = 0] = (\nu(0) - \mu(0))_+$, and $\mathbb{P}[X = 0, Y = 1] = (\mu(0) - \nu(0))_+$.

The following lemma gets us closer to our goal.

**LEM 9.22 (Coupling lemma)** Let $(X, Y)$ be any coupling of $\mu, \nu \in \Delta(S)$. Then

$$\|\mu - \nu\|_{TV} \leq \mathbb{P}[X \neq Y].$$

**Proof:** Note

$$\mu(s) = \mathbb{P}[X = s] = \mathbb{P}[X = s, X \neq Y] + \mathbb{P}[X = s, Y = s] \leq \mathbb{P}[X = s, X \neq Y] + \mathbb{P}[Y = s] \leq \mathbb{P}[X = s, X \neq Y] + \nu(s).$$

Similarly

$$(\nu(s) - \mu(s))_+ \leq \mathbb{P}[Y = s, X \neq Y],$$

so

$$|\mu(s) - \nu(s)| \leq \mathbb{P}[X = s, X \neq Y] + \mathbb{P}[Y = s, X \neq Y].$$
Summing over $y$ gives the result. (We also give an optimal coupling. Note that)

$$1 = \sum_{\omega \in \Omega} [\mu(\omega) \land \nu(\omega) + (\mu(\omega) - \nu(\omega))^+] = \sum_{\omega \in \Omega} [\mu(\omega) \land \nu(\omega) + (\nu(\omega) - \mu(\omega))^+],$$

so that

$$\sum_{\omega \in \Omega} \mu(\omega) \land \nu(\omega) = 1 - \|\mu - \nu\|_{TV}.$$  

Consider the following sub-intervals of $(0, 1)$. Divide up $(0, 1 - \|\mu - \nu\|_{TV})$ into disjoint intervals $I_\omega$ of length $\mu(\omega) \land \nu(\omega)$. Similarly, divide up $(1 - \|\mu - \nu\|_{TV}, 1)$ into disjoint intervals $J_\omega$ (respectively $K_\omega$) of length $\mu(\omega)$ (respectively $\nu(\omega)$). Then a coupling achieving $\|\mu - \nu\|_{TV} = \mathbb{P}[X \neq Y]$ is obtained by picking $U$ uniformly at random from $(0, 1)$ and assigning $X = Y = \omega$ if $U \in I_\omega$, or $X = \omega_1, Y = \omega_2$ if $U \in J_{\omega_1} \cap K_{\omega_2}$.)

We come back to the proof of the theorem.

**Proof:** By the coupling lemma, it suffices to find a coupling with high agreement probability. For each $1 \leq m \leq n$, we define

$$\mathbb{P}[X_{n,m} = x, Y_{n,m} = y] = \begin{cases} 
1 - p_{n,m} & \text{if } x = y = 0, \\
e^{-p_{n,m} - 1 + p_{n,m}} & \text{if } x = 1, y = 0, \\
e^{-p_{n,m} \frac{p_{n,m}^y}{y!}} & \text{if } x = 1, y \geq 1.
\end{cases}$$

The marginal of $X_{n,m}$ is Bernoulli with parameter $p_{n,m}$ and the marginal of $Y_{n,m}$ is Poisson with parameter $p_{n,m}$. (The goal is to make them as close as possible in distribution.) Therefore

$$Z =_d T_n = \sum_{1 \leq m \leq n} Y_{n,m} \sim \text{Poi}(\lambda).$$
We compute the disagreement probability. Note
\[
P[S_n \neq T_n] \leq \sum_{m \leq n} P[X_n,m \neq Y_n,m]
\]
\[
= \sum_{m \leq n} [e^{-p_{n,m}} - 1 + p_{n,m} + P[Y_n,m \geq 2]]
\]
\[
= \sum_{m \leq n} [e^{-p_{n,m}} + p_{n,m} - P[Y_n,m \leq 1]]
\]
\[
= \sum_{m \leq n} [e^{-p_{n,m}} + p_{n,m} - e^{-p_{n,m}} - p_{n,m}e^{-p_{n,m}}]
\]
\[
= \sum_{m \leq n} p_{n,m}[1 - e^{-p_{n,m}}]
\]
\[
\leq \sum_{m \leq n} p_{n,m}^2.
\]

\[
\text{EX 9.23 (Poisson approximation to the binomial)} \quad \text{Assume } p_{n,m} = \lambda/n \text{ for all } m. \text{ Then}
\]
\[
\|\text{Bin}(n, \lambda/n) - \text{Poi}(\lambda)\|_{TV} \leq \frac{\lambda^2}{n}.
\]

4.3 Example with dependence

\[
\text{EX 9.24 (Matching)} \quad \text{Let } S_n = \sum_{m=1}^n X_{n,m} \text{ be the number of fixed points in a}
\]
\[
\text{uniform random permutation, where } X_{n,m} = 1 \text{ if } m \text{ is a fixed point. We want to}
\]
\[
\text{compute } P[S_n = k]. \text{ Note that we cannot apply the previous theorem because of}
\]
\[
\text{the lack of independence. However, a Poisson limit with } \lambda = 1 \text{ seems natural. We}
\]
\[
\text{will need the following lemma.}
\]

\[
\text{LEM 9.25 (Inclusion-exclusion formula)} \quad \text{Let } A_1, A_2, \ldots, A_n \text{ be events and } A = \cup_{i=1}^n A_i. \text{ Then}
\]
\[
P[A] = \sum_{i=1}^n P[A_i] - \sum_{i<j} P[A_i \cap A_j]
\]
\[
+ \sum_{i<j<k} P[A_i \cap A_j \cap A_k] - \cdots + (-1)^{n-1}P[\cap_{i=1}^n A_i].
\]

(Moreover, truncating the sum at any term gives an upper bound if the next term is
negative and a lower bound if the next term is positive.)
Proof: Expand $1_A = 1 - \prod_{i=1}^n (1 - 1_{A_i})$ and take expectation. See [D].

Let $A_m = \{X_{n,m} = 1\}$. Then

$$\mathbb{P}[A] = n \frac{(n-1)!}{n!} - \frac{(n-2)!}{n!} \left(\frac{n}{2}\right) - \frac{(n-3)!}{n!} \left(\frac{n}{3}\right) - \cdots$$

So

$$P[S_n > 0] = \sum_{m=1}^{n} \frac{(-1)^{m-1}}{m!},$$

and

$$P[S_n = 0] = \sum_{m=0}^{n} \frac{(-1)^m}{m!}.$$

Note that the first two terms cancel each other out. Hence

$$\left|P[S_n = 0] - e^{-1}\right| = \left|\sum_{m=n+1}^{+\infty} \frac{(-1)^m}{m!}\right| \leq \frac{1}{(n+1)!} \left|\sum_{k=0}^{\infty} \frac{1}{(n+2)^k}\right|$$

$$= \frac{1}{(n+1)!} \left(1 - \frac{1}{n+2}\right)^{-1}.$$

Finally,

$$\mathbb{P}[S_n = k] = \binom{n}{k} \frac{1}{n(n-1) \cdots (n-k+1)} \mathbb{P}[S_{n-k} = 0]$$

$$= \frac{1}{k!} \mathbb{P}[S_{n-k} = 0]$$

$$\rightarrow e^{-1}.$$

References