6.262 Discrete Stochastic Processes, Spring 2009
Problem Set 2 — Solutions
due: Wednesday, February 18, 2009

Problem 1 (Exercise 1.22)

a Since X_i are independent, $\text{var}(S_n) = \text{var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{var}(X_i) \le nA$. By Chebyshev's inequality, it then follows that

$$\text{P}\left(\frac{|S_n - \text{E}(S_n)|}{n} > \epsilon\right) = \text{P}\left(|S_n - E(S_n)| > n\epsilon\right) \le \frac{\text{var}(S_n)}{n^2\epsilon^2} \le \frac{A}{n^2\epsilon^2}.$$

Thus, $\lim_{n\to\infty} \text{P}\left(\frac{|S_n - E(S_n)|}{n} > \epsilon\right) = 0$.

b Now, $\text{var}(S_n) = \text{var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{var}(X_i) \le \sum_{i=1}^n Ai^{1-\alpha} \le A + A\int_1^n x^{1-\alpha} = A + A(2-\alpha)^{-1}n^{2-\alpha}$. Thus,

$$\text{P}\left(\frac{|S_n - \text{E}(S_n)|}{n} > \epsilon\right) \le \frac{\text{var}(S_n)}{n\epsilon} \le \frac{A + A(2-\alpha)^{-1}n^{2-\alpha}}{n^2\epsilon^2} = \frac{A}{n^2\epsilon^2} + \frac{A}{n^\alpha(2-\alpha)\epsilon^2}.$$

As $\alpha > 0$, $\lim_{n\to\infty} \text{P}\left(\frac{|S_n - \text{E}(S_n)|}{n} > \epsilon\right) = 0$.

Problem 2

a Recall that

$$\text{E}(S_n^4) = \sum_{(i,j,k,m)\in\{1,2,\ldots,n\}^4} \text{E}(X_iX_jX_kX_m),$$

where the sum is taken over all 4-tuples taking values in the set $\{1, 2, \ldots, n\}$. Since X_i are independent, if $i \ne j$, then $\text{E}(X_iX_j) = \text{E}(X_i)\text{E}(X_j) = 0$, since X_i are zero-mean. It follows that the non-zero terms of the summation are of the form $i = j = k = m$ or $i = j, k = m$ or $i = k, j = m$ or $i = m, k = j$. There are n ways of chosing one value out of n, so there are n terms corresponding to $i = j = k = m$. Similarly, there are $n(n-1)$ terms corresponding to $i = j, k = m$, as well as $n(n-1)$ ways corresponding to $i = k, j = m$, and $n(n-1)$ ways corresponding to $i = m, k = j$. Also note that X_i are identically distributed, so $\text{E}(X_1^4) = \text{E}(X_i^4)$ and similarly for the other terms. Thus,

$$\text{E}(S_n^4) = n\text{E}(X_1^4) + 2n(n-1)\text{E}(X_1^2X_2^2) = n\gamma + 2n(n-1)\sigma^2.$$

b Noting that $E(S_n) = 0$ and, thus, $\text{var}(S_n) = \text{E}(S_n^4)$, we have

$$\begin{aligned} P(|S_n/n| > \epsilon) = P(|S_n| > n\epsilon) = P(|S_n - \text{E}(S_n)| > n\epsilon) = P(|S_n - \text{E}(S_n)|^4 > \epsilon^4 n^4) \\ \le \frac{\text{E}(|S_n - \text{E}(S_n)|^4)}{\epsilon^4 n^4} = \frac{n\gamma + 2n(n-1)\sigma^2}{\epsilon^4 n^4}, \end{aligned}$$

where the inequality follows by Markov.

c We now have

$$\sum_n P(|S_n/n| > \epsilon) \le \sum_n \frac{n\gamma + 2n(n-1)\sigma^2}{\epsilon^4 n^4} \le \sum_n \frac{\gamma}{\epsilon^4 n^3} + \sum_n \frac{2\sigma^2}{\epsilon^4 n^2} < \infty.$$

d Following the hint, it suffices to show that $\mathrm{E}(Y^2) < \infty \implies \mathrm{E}(Y) < \infty$ for Y a non-negative random variable. For that, notice

$$\mathrm{E}(Y) = \int_0^\infty y \, \mathrm{d}F(y) \le \int_0^1 y \, \mathrm{d}F(y) + \int_1^\infty y \, \mathrm{d}F(y)$$

where

$$\int_1^\infty y \, \mathrm{d}F(y) \le \int_1^\infty y^2 \, \mathrm{d}F(y) \le \mathrm{E}(Y^2) \quad \text{and} \quad \int_0^1 y \, \mathrm{d}F(y) \le \int_0^1 \mathrm{d}F(y) \le 1.$$

Therefore, $\mathrm{E}(Y) < \infty$.

e Suppose X_i no longer have zero mean. Let $W_n = S_n - \mathrm{E}(S_n)$. We still have that $\mathrm{E}(W_n^2)$ and $\mathrm{E}(W_n)$ are finite (why?), but the slightly tricky, though intuitively obvious, part is showing that $\mathrm{E}(S_n^4) < \infty \implies \mathrm{E}(W_n^4) < \infty$. Recall Jensen's inequality (Recitation, Feb. 17) where for ϕ convex,

$$\phi\left(\int_{\mathbb{R}} g(y)\mathrm{d}F_Y(y)\right) \le \int_{\mathbb{R}} \phi(g(y))\mathrm{d}F_Y(y).$$

Since x^p is convex for $p > 1$ and $x \ge 0$ (Hint: take its second derivative and check whether it's non-negative.), the above implies that $\mathrm{E}(|S_n|^p) \le \mathrm{E}(|S_n|^4) < \infty$ for all $0 < p \le 4$. Thus, $\mathrm{E}(|S_n|^4) < \infty \implies \mathrm{E}(|S_n|^3) < \infty$. But $\mathrm{E}(|S_n|^3) < \infty \implies \mathrm{E}(S_n^3) < \infty$. And we now see that $\mathrm{E}(W_n^4) = \mathrm{E}((S_n - \mathrm{E}(S_n))^4) = \mathrm{E}(S_n^3) + 3\mathrm{E}(S_n^2)\mathrm{E}(S_n) + 3\mathrm{E}(S_n)(\mathrm{E}(S_n))^2 + (\mathrm{E}(S_n))^3 < \infty$.

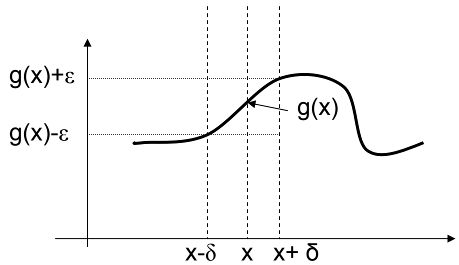
Now let $\mathrm{E}(W_n^4) = \gamma$ and $\mathrm{E}(W_n^2) = \sigma^2$. Then, $\mathrm{E}(W_n) = 0$ and $\mathrm{var}(W_n) = n\gamma + 2n(n-1)\sigma^2$ by the previous result. Furthermore,

$$P\left(\left|\frac{S_n - \mathrm{E}(S_n)}{n}\right| > \epsilon\right) = P(|S_n - \mathrm{E}(S_n)| > n\epsilon) = P(|W_n| > n\epsilon) = P(|W_n|^4 > \epsilon^4 n^4)$$
$$\le \frac{\mathrm{E}(|S_n - \mathrm{E}(S_n)|^4)}{\epsilon^4 n^4} = \frac{n\gamma + 2n(n-1)\sigma^2}{\epsilon^4 n^4}.$$

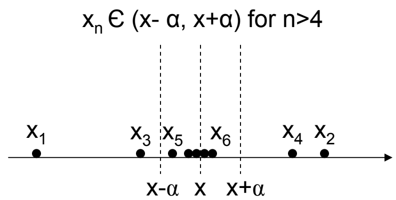
In that case, $\sum_n P\left(\left|\frac{S_n - \mathrm{E}(S_n)}{n}\right| > \epsilon\right) < \infty$, which, through the Borel-Cantelli Lemma, implies that $\dfrac{S_n - \mathrm{E}(S_n)}{n} \to 0$ with probability 1.

Problem 3

a



Continuity of g at x



Convergence of {x_n} to x

b Knowing when we're allowed to move the limit inside a function is one of the most useful things an elementary course on analysis teaches us. The following result, which will come in handy in the Feb. 23 recitation, is surprisingly simple to derive.

We want to show that if $x_n \to x$, then $g(x_n) \to g(x)$ if g is continuous. Recall that $x_n \to x$ iff for every $\alpha > 0$ there exists some integer N, such that $|x_n - x| < \alpha$ whenever $n > N$. Similarly, $g(x_n) \to g(x)$ if for every $n > 0$ there exists some integer M, such that $|g(x_n) - g(x)| < \nu$ whenever $n > M$.

Fix ν and notice that, since g is continuous, there exists some $\delta > 0$, such that $|g(x_n) - g(x)| < \nu$ whenever $|x_n - x| < \delta$. In turn, since $x_n \to x$, there exists some integer N such that $|x_n - x| < \delta$ for all $n > N$. It follows that for every $\nu > 0$, we can find some integer N such that $|g(x_n) - g(x)| < \nu$ whenever $n > N$. In other words $g(x_n) \to g(x)$.

As an example of how the theorem fails when g is not continuous, it suffices to let $g(x) = 0$ for $x < 0$ and $g(x) = 1$ for $x \geq 1$. Then for any sequence $x_n \to 0$ such that $x_n < 0$ for all n, we have that $g(x_n) \to 0 \neq g(0)$.

c Now let X_n be a sequence of random variables converging to some $x \in \mathbb{R}$ with probability 1 (this is frequently denoted as $X_n \to x$ a.s. or $X_n \to x$ a.e.). Given a fixed a realization of a random variable, the problem reduces to its deterministic version. Specifically, for all realizations ω for which $X_n(\omega) \to x$, part a) tells us that $g(X_n(\omega)) \to g(x)$. It follows that

$$\{\text{realizations } \omega \mid X_n(\omega) \to x\} \subset \{\text{realizations } \omega \mid g(X_n(\omega)) \to g(x)\},$$

and thus, $\mathrm{P}(\{\text{realizations } \omega \mid X_n(\omega) \to x\}) \leq \mathrm{P}(\{\text{realizations } \omega \mid g(X_n(\omega)) \to g(x)\})$. Since $\mathrm{P}(\{\text{realizations } \omega \mid X_n(\omega) \to x\}) = 1$, we obtain $\mathrm{P}(\{\text{realizations } \omega \mid g(X_n(\omega)) \to g(x)\}) = 1$, or $g(X_n) \to g(x)$ a.s.

d What about if $X_n \to x$ in probability? We know it's a weaker form of convergence, but does it still imply that $g(X_n) \to g(X)$ in probability? Going back to the deterministic case, we know that for every $\epsilon > 0$, there exists a $\delta > 0$ such that $|x_n - x| < \delta \implies |g(x_n) - g(x)| < \epsilon$.

This means that for every $\epsilon > 0$, there is some $\delta > 0$ (the same δ will work), such that $|g(x_n) - g(x)| > \epsilon \implies |x_n - x| > \delta$. It follows that

$$\{\text{realizations } \omega \mid |g(X_n(\omega) - g(x)| > \epsilon\} \subset \{\text{realizations } \omega \mid |X_n(\omega) - x| > \delta\},$$

and so,

$$\text{P}(\text{realizations } \omega \mid |g(X_n(\omega) - g(x)| > \epsilon) \leq \text{P}(\text{realizations } \omega \mid |X_n(\omega) - x| > \delta).$$

Since $\lim_{n\to\infty} \text{P}(\text{realizations } \omega \mid |X_n(\omega) - x| > \delta) = 0$ for every $\delta > 0$ (definition of convergence in probability), it follows that for every $\epsilon > 0$,

$$\lim_{n\to\infty} \text{P}(\text{realizations } \omega \mid |g(X_n(\omega) - g(x)| > \epsilon) = 0,$$

that is, $g(X_n) \to g(x)$ in probability. Note: can you see why this proof relies on X being degenerate, i.e. $X(\omega) = x$ for all ω?

e Now we have that $X_n \to x$, for some $x \in \mathbb{R}$, in distribution. While the convergence in distribution does generally not imply convergence in probability (recall the example mentioned at the end of February 19 lecture), the implication holds true if X_n converges to a constant, and that's what we will show presently. Consider the random variable X that equals x with probability 1. Note that the corresponding CDF is given by $F_X(\alpha) = 1$ if $\alpha \geq x$ and $F_X(\alpha) = 0$ if $\alpha < x$. Now fix $n \in \mathbb{Z}_+$ and let $\epsilon > 0$. Then $\text{P}(|X_n - x| \leq \epsilon) = \text{P}(-\epsilon \leq X_n - x \leq \epsilon) = F_{X_n}((-\epsilon)^-) - F_{X_n}(\epsilon)$. But, since $X_n \to x$ in distribution, then $F_{X_n}(\alpha) \to F_X(\alpha)$ for all $\alpha \in \mathbb{R}$. So, $\lim_{n\to\infty} \text{P}(|X_n - x| \leq \epsilon) = \lim_{n\to\infty}(F_{X_n}(x + \epsilon) - F_{X_n}((x - \epsilon)^-)) = \lim_{n\to\infty} F_{X_n}(x + \epsilon) - \lim_{n\to\infty} F_{X_n}((x - \epsilon)^-) = 1 - 0 = 1$. Thus, $\lim_{n\to\infty} \text{P}(|X_n - x| > \epsilon) = 0$ and the result follows. Note:

- We were allowed to distribute the limit across the two terms because both individual limits were finite.
- Can you see why this proof relies on X being degenerate, i.e. $X(\omega) = x$ for all ω?

Now that we know that $X_n \to x$ in probability, then $g(X_n) \to g(x)$ in probability by part d). Finally, since convergence in probability implies convergence in distribution, we have $g(X_n) \to g(x)$ in distribution.

Note that parts c) through e) hold for $X_n \to X$, where X need not be a trivial random variable. However, some of the proof techniques in parts d) and e) no longer apply (try to identify them!). The more general proofs are available in one of the reference texts for the course (Grimmet and Stirzaker, used in 6.436).

f Finally, we have that $X_n \to x$ in mean square, that is, $\text{E}(|X_n - x|^2) \to 0$ as $n \to \infty$. As an example of such convergence, consider the sequence of random variables X_n, where $\text{P}(X_n = n) = \text{P}(X_n = -n) = 1/(2n^3)$ and $\text{P}(X_n = 0) = 1 - 1/n^3$. Then, $\text{E}(X_n) = 0$ and $\text{E}(|X_n - 0|^2) = n^2/n^3 = 1/n \to 0$, so $X_n \to 0$ in mean square. (By the way, does $X_n \to 0$ in any other sense?) Consider a continuous function $g(y) = y^3$. Then, $\text{P}(g(X_n) = n^3) = \text{P}(g(X_n) = -n^3) = 1/(2n^3)$ and $\text{P}(X_n = 0) = 1 - 1/n^3$. Thus, $\text{E}(g(X_n)) = 0$, but $\text{E}(|g(X_n) - 0|^2) = n^6/n^3 = n^3 \to \infty$, so $g(X_n)$ does not converge to $g(x) = 0$ in mean square.

Problem 4 (**Exercise 2.1**)

We want to show that $f_{S_n}(t) = \dfrac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$.

a We proceed by induction. For $n = 1$, $f_{S_1}(t) = \lambda e_{-\lambda t}$ and the expression holds. Suppose that the expression holds for n. Since $S_n = S_n + X_{n+1}$ and S_n is independent of X_{n+1}, we have

$$f_{S_{n+1}}(t) = f_{S_n}(t) * f_{X_{n+1}}(t) = \int_0^t \frac{\lambda^n e^{-\lambda x} x^{n-1}}{(n-1)!} \lambda e^{-\lambda(t-x)} \mathrm{d}x = \frac{\lambda^{n+1} t^n e^{-\lambda t}}{n!}.$$

b Since $g_{X_i}(r) = \dfrac{\lambda}{\lambda - r}$ for $r < \lambda$, we have

$$g_{S_n}(r) = \mathrm{E}(e^{rS_n}) = \mathrm{E}(e^{r \sum_{i=1}^n X_i}) = \mathrm{E}\left(\prod_{i=1}^n e^{rX_i}\right) = \prod_{i=1}^n \mathrm{E}(e^{rX_i}) = \left(\frac{\lambda}{\lambda - r}\right)^n \quad \text{for } r < \lambda,$$

where in order to distribute the expectation across the product, we used the fact that X_i are independent. To obtain the inverse transform, note that

$$g_{S_n}(r) = \frac{\lambda^{n-1}}{(n-1)!} \frac{\mathrm{d}^n}{\mathrm{d}r^n} g_X(r).$$

Recalling that $g_X(r) = \int_0^\infty e^{rx} \lambda e^{-\lambda x} \mathrm{d}x$ and taking the derivative inside of the integral[1],

$$\frac{\lambda^{n-1}}{(n-1)!} \frac{\mathrm{d}^n}{\mathrm{d}r^n} g_X(r) = \int_0^\infty \frac{\lambda^n e^{-\lambda x} x^{n-1}}{(n-1)!} e^{rx} \mathrm{d}x,$$

and so, matching terms, $f_{S_n}(x) = \dfrac{\lambda^n e^{-\lambda x} x^{n-1}}{(n-1)!}$.

c From class, $f_{S_1,\ldots,S_n}(s_1, \ldots, s_n) = \lambda^n e^{-\lambda s_n}$ for $0 \le s_1 \le \ldots \le s_n$. Note that $f_{S_2,\ldots,S_n}(s_2, \ldots, s_n) = \int_0^{s_2} f_{S_1,\ldots,S_n}(x, s_2 \ldots, s_n) \mathrm{d}x = s_2 \lambda^n e^{-\lambda s_n}$. Similarly, $f_{S_3,\ldots,S_n}(s_3, \ldots, s_n) = \int_0^{s_2} f_{S_2,\ldots,S_n}(x, s_3 \ldots, s_n) \mathrm{d}x = \int_0^{s_2} x \lambda^n e^{-\lambda s_n} \mathrm{d}x = \frac{x^2}{2} \lambda^n e^{-\lambda s_n}$. Pursuing the process in order to eliminate $S_1, \ldots, S_{n-1}$, we reach $f_{S_n}(t) = \dfrac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$.

Problem 5 (**Exercise 2.3**)

a Given $S_n = \tau$, we have that $N(t) = n$ for $\tau < t$ iff there are no arivals from τ to t. Then, $\mathrm{P}(N(t) = n \mid S_n = \tau) = \mathrm{P}(\text{no arrivals in } t - \tau \mid S_n = \tau) = \mathrm{P}(\text{no arrivals in } t - \tau)$, where the last equality follows by the memoryless property of the Poisson process. Thus,

$$\mathrm{P}(N(t) = n \mid S_n = \tau) = e^{-\lambda(t-\tau)}.$$

[1] This is sometimes a dangerous thing to do, but we're fine here, since, loosely speaking, the integrand is well-behaved. A more formal way to do this could, for instance, consist of showing that the derivative, viewed as a limit of a difference quotient, approaches the result monotonically.

b Since $P(N(t) = n) = \int_{\tau=0}^{t} \mathrm{P}(N(t) = n | S_n = \tau) f_{S_n}(\tau) \mathrm{d}\tau$, applying the result of a) yields

$$\begin{aligned} P(N(t) = n) &= \int_{\tau=0}^{t} e^{-\lambda(t-\tau)} \frac{\lambda^n \tau^{n-1} e^{-\lambda\tau}}{(n-1)!} \mathrm{d}\tau = \int_{\tau=0}^{t} \frac{\lambda^n \tau^{n-1} e^{-\lambda t}}{(n-1)!} \mathrm{d}\tau = \frac{\lambda^n e^{-\lambda t}}{(n-1)!} \int_{\tau=0}^{t} \tau^{n-1} \mathrm{d}\tau \\ &= \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \end{aligned}$$

which is the desired result.

Problem 6

a For $t = 1, 2, \ldots$, let $\{Y_t\}$ denote the sequence of outcomes in a Bernoulli process, where $Y_t = 1$ corresponds to a "success" (or arrival) occurring on the t^{th} trial and $\mathrm{P}(Y_t = 1) = q$. Recall that the corresponding number of successes in t trials is given by $N(t)$, which is binomially distributed. Adopting this view of the binomial distribution, it becomes easy to see that

$$\begin{aligned} p_{N(t+1)}(k) &= \mathrm{P}(\{N(t) = k \cap X_t = 0\} \cup \{N(t) = k-1 \cap X_t = 1\}) \\ &= \mathrm{P}(\{N(t) = k \cap Y_{t+1} = 0\}) + \mathrm{P}(\{N(t) = k-1 \cap Y_{t+1} = 1\}) \quad \text{disjoint events} \\ &= \mathrm{P}(N(t) = k)\mathrm{P}(Y_{t+1} = 0) + \mathrm{P}(N(t-1) = k-1)\mathrm{P}(Y_{t+1} = 1) \quad \text{Bernoulli proc. is memoryless} \\ &= \mathrm{P}(N(t) = k)(1-q) + \mathrm{P}(N(t-1) = k-1)q. \end{aligned}$$

Note that a somewhat more useful (in the context of part c) form of this result is

$$p_{N(t+1)}(k) = \mathrm{P}(Y_{t+1} = 0 | N(t) = k) p_{N(t)}(k) + \mathrm{P}(Y_{t+1} = 1 | N(t) = k) p_{N(t)}(k-1).$$

b Following a similar reasoning, $p_{N(t+1)}(0) = \mathrm{P}(Y_1 = 0, Y_2 = 0, \ldots, Y_{t+1} = 0) = (1-q)^{t+1}$. By the Principle of Recursive Definition, the solution to the above recursion is unique. Since the recursion is satisfied by the binomial distribution for $t \geq 1$ and $k \geq 0$, it follows that the recursion is an alternative way to specify the binomial distribution.

c Let $q = 1/2$. We wish to change the distribution on Y_1, Y_2, Y_3 so that the distribution of $N(t)$ remains binomial, but the process is no longer Bernoulli.

Parts a) and b) provide an easy criterion to ensure the binomial distribution on $N(t)$. Starting with $t = 1$, $p_{N(1)}(0) = p_{N(t)}(0) = 1/2$. It follows that of all the possible $\{0,1\}^3$ realizations of (Y_1, Y_2, Y_3), exactly half of them have $Y_1 = 0$. Thus, populating the first row in the table of possible realizations yields

00001111.

Continuing on, $p_{N(2)}(0) = p_{N(1)}(0)1/2$. In other words, conditioned on $N(1) = 0$, exactly half of the possible (Y_1, Y_2, Y_3) triples will have $X_2 = 0$. Partially populating the second row in the table of possible realizations yields

0011

00001111.

Still at $t = 2$, $p_{N(2)}(1) = p_{N(1)}(0)1/2 + p_{N(1)}(1)1/2$. Thus, conditioned on $N(1) = 1$, exactly half of the possible (Y_1, Y_2, Y_3) triples will have $X_2 = 0$. Further populating the second row in the table of possible realizations yields

$$\begin{array}{l} 00110011 \\ 00001111. \end{array}$$

Note that $p_{N(2)}(2) = p_{N(1)}(1)1/2$ is consistent with the above list, as given $X_1 = 1$, half of the patterns correspond to $X_2 = 0$. So far, all of the possible patterns are equally represented. The choice occurs when looking at $t = 3$. Specifically, $p_{N(3)}(0) = p_{N(2)}(0)1/2$, which yields

$$\begin{array}{l} 01 \\ 00110011 \\ 00001111. \end{array}$$

Similarly, $p_{N(3)}(3) = p_{N(2)}(2)1/2$ and so the possible patterns become

$$\begin{array}{l} 01 \quad\;\; 01 \\ 00110011 \\ 00001111. \end{array}$$

Finally, $p_{N(3)}(2) = p_{N(2)}(2)1/2 + p_{N(2)}(1)1/2 = (p_{N(2)}(2) + p_{N(2)}(1))1/2$, which only tells us that conditioned on $\{(X_1, X_2) = (0, 1) \cup \{(X_1, X_2) = (1, 0)\}$, $X_3 = 0$ with equal probability. Among the $\binom{4}{2}$ resulting ways of distributing the two $1'$s and two $0'$s in the remaining slots, four yield the complete set $\{0, 1\}^3$. The remaining two, namely,

$$\begin{array}{l} 01001101 \\ 00110011 \\ 00001111 \end{array}$$

and

$$\begin{array}{l} 01110001 \\ 00110011 \\ 00001111 \end{array}$$

yield a probability distribution for $N(t)$ that is binomial, but do not specify a Bernoulli process (otherwise, all 8 patterns would be present) even if we let Y_i for $i > 3$ be IID Bernoulli random variables.