6.262 Discrete Stochastic Processes, Spring 2009
Problem Set 1 — Solutions
due: Wednesday, February 11, 2009

Problem 1

a (**Exercise 1.12)**
Note that X_n is a local minimum if it is the smallest of the 3 IID random variables X_{n-1}, X_n, and X_{n+1}. One is strictly smaller than the others with probability 1 because the variables are continuous (i.e. the probability that any two are equal is zero). By symmetry, each variable is equally likely to be the smallest, so

$$P(X_n \le X_{n+1}, X_n \le X_{n-1}) = 1/3.$$

b The answer is very similar to that of the previous problem, but let's be a little more systematic. With probability 1, the random variables $X_1, X_2, \ldots, X_n$ can be ordered to form a strictly increasing finite sequence since the random variables are continuous. It follows that $(X_1, X_2, \ldots, X_n)$ can be put in correspondence with all possible orderings of the set $\{1, 2, \ldots, n\}$. For instance, for $n = 3$, the event $\{X_3 > X_1 > X_2\}$ then corresponds to $(2, 1, 3)$. Now it's easy: since the random variables are IID, all the $n!$ orderings of $\{1, 2, \ldots, n\}$ are equally likely. Since the events of interest correspond to having n in the n^{th} position, it follows that the desired probability is $(n-1)!/n! = 1/n$.

c Again, let's be systematic. Let I_n denote the indicator associated with the event that X_n is the record to date. Then, $\mathrm{E}[\sum_{n=1}^{\infty} I_n]$ yields the expected number of records in the sequences $X_1, X_2, \ldots$. Since expectation is linear, the quantity of interest becomes, $\sum_{n=1}^{\infty} \mathrm{E}[I_n] = \sum_{i=1}^{\infty} n^{-1} = \infty$.

d Let $R_n = \{X_{n+1}$ is a record to date $\}$. Writing out the conditional probability,

$$\mathrm{P}(R_{n+1} \mid R_n) = \frac{\mathrm{P}(R_{n+1} \cap R_n)}{\mathrm{P}(R_n)}.$$

Reasoning similarly as in b), $\mathrm{P}(R_{n+1} \cap R_n) = (n-1)!/(n+1)! = \frac{1}{n(n+1)}$, as we now need n in the n^{th} position and $n+1$ in the $(n+1)^{\text{st}}$ position. Also, $\mathrm{P}(R_n) = 1/n$ as found in b). Thus, $\mathrm{P}(R_{n+1} \mid R_n) = 1/(n+1) = \mathrm{P}(R_{n+1})$ and the events are independent!

Can you see how to obtain the same answer by counting sequences? Hint: the conditional universe contains all sequences of length $n+1$ with n in either the n^{th} or $(n+1)^{\text{st}}$ position. (Thanks to Sam for making the TA aware of this.)

e Now let I_n denote the indicator associated with the event that both X_n and X_{n+1} are records to date. As previously, $\mathrm{P}(R_{n+1} \cap R_n) = \frac{1}{n(n+1)}$. Thus, $\mathrm{E}[I_n] = \frac{1}{n(n+1)}$. Since $\mathrm{E}[\sum_{n=1}^{\infty} I_n]$ yields the expected number of adjacent pairs of records in the sequence $X_1, X_2, \ldots$, the desired result is $\sum_{n=1}^{\infty} \mathrm{E}[I_n] = \sum_{n=1}^{\infty} \frac{1}{n(n+1)} \le \sum_{n=1}^{\infty} \frac{1}{n^2} \le 1 + \int_1^{\infty} \frac{1}{x^2} dx < \infty$.

(Actually, $\sum_{i=1}^{\infty} \frac{1}{n^2}$ evaluates to $\zeta(2) = \pi^2/6$, where ζ denotes the Riemann zeta function. If you haven't heard of the latter, it suffices to say that it's a hugely interesting object in its own right. But, even without bringing ζ into the picture, there are a gazillion ways to evaluate $\sum_{i=1}^{\infty} n^{-2}$— fourteen of those can be found at:
`http://www.secamlocal.ex.ac.uk/people/staff/rjchapma/etc/zeta2.pdf`)

Note: There were many ways of approaching this problem other than counting patterns. However, no matter how you end up doing it, the point is to be systematic about it — not just divide and add things ad hoc because it *seems* like the right thing to do.

Problem 2

a (**Exercise 1.16)**
We first find $\mathrm{P}(Z = n)$. The quickest way to do this, by far, is through transforms. Alternatively, in order to solve it as a convolution of probability distributions, it is useful to recall from the Binomial Theorem. In particular, recall that $(\lambda+\mu)^n = \sum_{i=0}^{n}\binom{n}{i}\lambda^i\mu^{n-i}$. Then, for $n \geq 0$, we have:

$$\begin{aligned}\mathrm{P}(Z=n) &= \sum_{i=0}^{n}\mathrm{P}(X=i)\mathrm{P}(Y=n-i)\\ &= \sum_{i=0}^{n}\left(\frac{\lambda^i e^{-\lambda}}{i!}\right)\left(\frac{\mu^{n-i}e^{-\mu}}{(n-i)!}\right)\\ &= \frac{e^{-(\lambda+\mu)}}{n!}\sum_{i=0}^{n}\binom{n}{i}\lambda^i\mu^{n-i}\\ &= \frac{(\lambda+\mu)^n e^{-(\lambda+\mu)}}{n!}\end{aligned}$$

We have just shown that the sum of independent Poisson PMFs is a Poisson PMF with new rate equal to the sum of the individual rates.

For $0 \leq m \leq n$, we have:

$$\begin{aligned}P(Y=m|Z=n) &= \frac{P(Z=n|Y=m)P(Y=m)}{P(Z=n)}\\ &= \left(\frac{\lambda^{n-m}e^{-\lambda}}{(n-m)!}\right)\left(\frac{\mu^m e^{-\mu}}{m!}\right)\left(\frac{(\lambda+\mu)^n e^{-(\lambda+\mu)}}{n!}\right)^{-1}\\ &= \binom{n}{m}\left(\frac{\lambda}{\lambda+\mu}\right)^{n-m}\left(\frac{\mu}{\lambda+\mu}\right)^m\end{aligned}$$

This is a binomial PMF and we will see later on in Chapter 2 why this is to be expected.

b Let $Y_1, Y_2, \ldots$ be a sequence of IID Poisson r.v's with mean 1 (hence the variance is also 1). Denote the sum as $S_n = Y_1 + \ldots + Y_n$ then we know that adding poisson r.v's gives a poisson r.v and hence S_n is poisson with mean n. From the central limit theorem we have,

$$\lim_{n\to\infty} P\left(\frac{S_n - n}{\sqrt{n}} \leq 1\right) = \int_{-\infty}^{1}\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx \approx 0.84$$

Since S_n and X_n have the same distribution we have $P\left(X_n \leq n + \sqrt{n}\right) = P\left(\frac{X_n - n}{\sqrt{n}} \leq 1\right) = P\left(\frac{S_n-n}{\sqrt{n}} \leq 1\right)$ and the result follows from above.

Problem 3

a We are interested in calculating the probability that S_n is contained within an interval of fixed width $2m+1$ centered at the mean, $n\delta$. As n increases, the standard deviation of S_n increases as well (as $\sqrt{n}$, in fact), and we expect the probability mass associated with the interval to vanish in the limit.

Intuition is useful, but intuition can be wrong, so let's do this precisely. Let Y_n denote the zero-mean, unit-variance version of S_n, that is, $Y_n = (S_n - n\delta)/\sqrt{n}\sigma$. (Note that $\sigma = \sqrt{\text{var}(X_1)} = \sqrt{\delta(1-\delta)}$ does not vary with n.) Then,

$$\sum_{n\delta - m \leq i \leq n\delta + m} P(S_n = i) = F_{S_n}(n\delta + m) - F_{S_n}(n\delta - m)^- = F_{Y_n}\left(\frac{m}{\sqrt{n}\sigma}\right) - F_{Y_n}\left(-\frac{m}{\sqrt{n}\sigma}\right)^-$$

Note 1: Can you see why $\sum_{i=m_1}^{m_2} P(S_n = i) = F_{S_n}(m_2) - F_{S_n}(m_1)^- \neq F_{S_n}(m_2) - F_{S_n}(m_1)$ for S_n discrete?

For large n, the distribution of Y_n converges to that of a standard normal, so

$$\lim_{n\to\infty} \sum_{i:n\delta - m \leq i \leq n\delta + m} P(S_n = i) = \lim_{n\to\infty}\left(F_{Y_n}\left(\frac{m}{\sqrt{n}\sigma}\right) - F_{Y_n}\left(-\frac{m}{\sqrt{n}\sigma}\right)^-\right) = \frac{1}{2} - \frac{1}{2} = 0.$$

Note 2: Though this was not required, had we wished to be rigorous, we could have taken the above limit as follows. First note that

$$F_{Y_n}\left(\frac{m}{\sqrt{n}\sigma}\right) - F_{Y_n}\left(-\frac{m}{\sqrt{n}\sigma} - 1\right) \geq \sum_{n\delta - m \geq i \leq n\delta + m} P(S_n = i) \leq F_{Y_n}\left(\frac{m}{\sqrt{n}\sigma}\right) - F_{Y_n}\left(-\frac{m}{\sqrt{n}\sigma}\right).$$

(This way, we don't have to worry about Note 1. The trick works here as both the upper and the lower bound will converge to zero. This manner of "sandwiching" the argument is typically very useful—keep it in mind.) Now we need to show that $F_{Y_n}(y_n) \to \phi(y)$ (ϕ is the cumulative gaussian distribution), where the sequence $y_n \to y$ and the functions $F_{Y_n} \to \phi$ (pointwise). Consider an $\epsilon > 0$ and pick N large enough so that for all $n \geq N$ we have, $y - \epsilon \leq y_n \leq y + \epsilon$. Since F_{Y_n} is non-decreasing, we get $F_{Y_n}(y-\epsilon) \leq F_{Y_n}(y_n) \leq F_{Y_n}(y+\epsilon)$ for all $n \geq N$. Take the limit $n \to \infty$ which gives $\phi(y-\epsilon) \leq \lim_{n\to\infty} F_{Y_n}(y_n) \leq \phi(y+\epsilon)$ (by Central Limit Theorem). Now, take the limit $\epsilon \to 0$ and since $\phi(\cdot)$ is a continuous function, both sides of the inequality converge to the same value $\phi(y)$ which completes the proof.

Note 3: Though the intuitive explanation at the beginning of this solution is tempting and may pass off as an answer in other courses, in 6.262 we need to be precise. Intuition is helpful, but it is often wrong. As an illustration, consider a random variable Z_n such that $\text{P}(Z_n = -\sqrt{n}) = \text{P}(Z_n = 0) = \text{P}(Z_n = \sqrt{n}) = 1/3$. Here also, the standard deviation of Z_n increases as $\sqrt{n}$. Yet, any fixed interval about the mean (origin) will contain Z_n with probability $1/3$ for all sufficiently large n. The desired probability does therefore not vanish.

Another potentially misleading intuitive argument is that the interval of interest is sliding into the tails of the distribution with n, while the standard deviation increases only as $\sqrt{n}$ (and so, in the limit, the interval spans a probability mass of zero). As an illustration, consider the random variable Y_n such that $\text{P}(Y_n = \delta n) = 1/2 = P(Y_n = \delta n - 2\sqrt{n})$. Then, $\text{E}(Y_n) = \delta n - \sqrt{n}$ and $\text{var}(Y_n) = (\sqrt{n})^2/2 + (\sqrt{n})^2/2 =$

n. So, the standard deviation of Y_n increases as $\sqrt{n}$, the interval $[\delta n - m, \delta n + m]$ slides away as n, but $P(Y_n \in [\delta n - m, \delta n + m]) = 1/2$ for all n.

b Unlike part a), we now have an interval that begins at the origin, and grows to include more and more of the probability mass. The interval grows at the rate n, while the variance of the probability distribution grows as $\sqrt{n}$, so we may expect the interval to include exactly half of the probability mass in the limit.

We have:

$$\lim_{n\to\infty} \sum_{0 \le i \le n\delta + m} P(S_n = i) = \lim_{n\to\infty} \left(F_{Y_n}\left(\frac{m}{\sqrt{n}\sigma}\right) - F_{Y_n}\left(\frac{-n\delta}{\sqrt{n}\sigma}\right)^{-} \right) = \frac{1}{2} - 0 = \frac{1}{2}.$$

c Now the interval of summation is increasing with n in both directions, while the standard deviation is still increasing with $\sqrt{n}$. Thus, we may expect the interval to include all of the probability mass in the limit.

We have:

$$\lim_{n\to\infty} \sum_{n(\delta - 1/m) \le i \le n(\delta + 1/m)} P(S_n = i) = \lim_{n\to\infty} \left(F_{Y_n}\left(\frac{n}{m\sqrt{n}\sigma}\right) - F_{Y_n}\left(\frac{-n}{m\sqrt{n}\sigma}\right)^{-} \right) = 1 - 0 = 1.$$

Problem 4

a With the non-compounding strategy, $\Delta W_n = \sum_{i=1}^n X_i - n$. Since $\mathrm{E}(X_i) = (2)(0.5) + (0.25)(0.5) = 1.125$, we have that $\mathrm{E}(\Delta W_n) = \mathrm{E}(\sum_{i=1}^n X_i - n) = \sum_{i=1}^n \mathrm{E}(X_i) - n = n/8$. Also, $\mathrm{var}(\Delta W_n) = \mathrm{var}(\sum_{i=1}^n X_i - n) = \mathrm{var}(\sum_{i=1}^n X_i)$ and, since X_i are independent, $\mathrm{var}(\Delta W_n) = \sum_{i=1}^n \mathrm{var}(X_i) = 0.765625n$, where $\mathrm{var}(X_i) = \mathrm{E}(X_i^2) - E^2(X_i) = (2)^2(0.5) + (0.25)^2(0.5) - 1.125^2 = 0.765625$. Finally, $\sigma_{\Delta W_n} = \sqrt{\mathrm{var}(\Delta W_n)}$.

b With the compounding strategy, we have $\Delta \tilde{W}_n = \prod_{i=1}^{n} X_i - 1$. By the linearity of expectation and since X_i are still independent, $\mathrm{E}(\Delta \tilde{W}_n) = \prod_{i=1}^n \mathrm{E}(X_i) - 1 = 1.125^n - 1$. As we hope you learned in 6.041, variances generally don't factor (but can you think of a simple rule for when they actually do?). So, we proceed from the definition: $\mathrm{var}(\Delta \tilde{W}_n) = \mathrm{var}(\prod_{i=1}^n X_i - 1) = \mathrm{var}(\prod_{i=1}^n X_i) = \mathrm{E}(\prod_{i=1}^n X_i)^2 - \mathrm{E}^2(\prod_{i=1}^n X_i)$. Since X_i are independent, $\mathrm{E}(\prod_{i=1}^n X_i)^2 = \prod_{i=1}^n \mathrm{E}(X_i^2) = ((2)^2(0.5) + (0.25)^2(0.5))^n = 2.03125^n$ and $\mathrm{E}^2(\prod_{i=1}^n X_i) = 1.125^{2n}$, so $\mathrm{var}(\Delta \tilde{W}_n) = 2.03125^n - 1.125^{2n} = 2.03125^n - 1.265625^n$. Finally, $\sigma_{\Delta \tilde{W}_n} = \sqrt{\mathrm{var}(\Delta \tilde{W}_n)}$.

c $E(\Delta \tilde{W}_n) >> E(\Delta W_n)$.

Isn't the compounding strategy so much better?

d Let k and $\tilde{k}$ be the number of heads required to win or break even in the non-compounding and compounding strategies, respectively. Then, for $2k + 0.25(n-k) - n \geq 0$, we need $k \geq 0.75(n)/1.75 = 3n/7$. Similarly, for $2^{\tilde{k}} \times 0.25^{n-\tilde{k}} - 1 \geq 0$, we need $2^{\tilde{k}} 0.25^{-\tilde{k}} = 2^{3\tilde{k}} \geq 0.25^{-n} = 2^{2n}$, that is, $\tilde{k} \geq 2n/3$. So, with the compounding strategy, we have to be significantly luckier in order to win or break even. Perhaps compounding isn't such a great idea after all.

e Using the result in d) and letting K_{100} denote the number of heads in 100 tosses,

$$\mathrm{P}(\Delta W_n \geq 0) = \mathrm{P}(K_{100} \geq 43) = \mathrm{P}\left(\frac{K_{100} - E(K_{100})}{\sqrt{\mathrm{var}(K_{100})}} \geq \frac{43 - E(K_{100})}{\sqrt{\mathrm{var}(K_{100})}}\right),$$

where $\mathrm{E}(K_{100}) = 100/2 = 50$ and $\mathrm{var}(K_{100}) = 100/4 = 25$. Applying the Central Limit Theorem as an approximation,

$$\mathrm{P}\left(\frac{K_{100} - E(K_{100})}{\sqrt{\mathrm{var}(K_{100})}} \geq \frac{43 - E(K_{100})}{\sqrt{\mathrm{var}(K_{100})}}\right) \approx \Phi\left(\frac{42.5 - E(K_{100})}{\sqrt{\mathrm{var}(K_{100})}})\right) = 1 - \Phi(-1.4) = 0.9332.$$

Note that, calculated exactly to four significant digits, $\mathrm{P}(K_{100} \geq 43) = 0.9333$.

Similarly for the compounding strategy,

$$\mathrm{P}(\Delta \tilde{W}_n \geq 0) = \mathrm{P}(K_{100} \geq 67) = \mathrm{P}\left(\frac{K_{100} - E(K_{100})}{\sqrt{\mathrm{var}(K_{100})}} \geq \frac{67 - E(K_{100})}{\sqrt{\mathrm{var}(K_{100})}}\right),$$

where $\mathrm{E}(K_{100})$ and $\mathrm{var}(K_{100})$ are as before. Applying the Central Limit Theorem as an approximation,

$$\mathrm{P}\left(\frac{K_{100} - E(K_{100})}{\sqrt{\mathrm{var}(K_{100})}} \geq \frac{67 - E(K_{100})}{\sqrt{\mathrm{var}(K_{100})}}\right) \approx \Phi\left(\frac{66.5 - E(K_{100})}{\sqrt{\mathrm{var}(K_{100})}})\right) = 1 - \Phi(3.3) = 0.0004834.$$

Note that, calculated exactly to four significant digits, $\mathrm{P}(K_{100} \geq 43) = 0.0004369$.

f The compounding strategy has a much larger expected gain than the non-compounding strategy. That's roughly due to the fact that with the compounding strategy, we can acquire a large wealth with some non-zero probability (as much as $\$2^n - 1$, which is a decent sum for large n, compared to only $\$n$ with the non-compounding strategy). However, our chances of losing are drastically greater as well. So, the compounding strategy carries a small probability of a large gain, while the non-compounding strategy carries a large probability of a small to medium gain.

Focusing on risks rather than rewards yields another peculiar aspect of the compounding strategy, namely, that we can loose at most \$1 compared to $\$n$ with the non-compounding strategy. Suppose you decide to gamble \$1. Which strategy would you choose? Would you change your answer if, instead, you had \$10000?

g As calculated in d), to win or break even using the compounding strategy, we need more than 2(100)/3 heads. Thus, letting K_{100} denote the number of heads in 100 tosses, $\mathrm{P}(\Delta \tilde{W}_n \geq 0) =$

$\mathrm{P}(K_{100} \geq 67)$. But, note that

$$\begin{aligned}\mathrm{P}(K_{100} \geq 67) &= \mathrm{P}(K_{100} - \mathrm{E}(K_{100}) \geq 67 - \mathrm{E}(K_{100})) \\ &= \mathrm{P}(K_{100} - \mathrm{E}(K_{100}) \geq 17) \\ &= \frac{1}{2}\mathrm{P}(|K_{100} - \mathrm{E}(K_{100})| \geq 17),\end{aligned}$$

where $\mathrm{E}(K_{100}) = 100/2 = 50$ and $\mathrm{var}(K_{100}) = 100/4 = 25$ and the last equality follows from the fact that K_{100} is symmetric about its mean. Applying the Chebyshev bound to this probability yields

$$\frac{1}{2}\mathrm{P}(|K_{100} - \mathrm{E}(K_{100})| \geq 17) \leq \frac{\mathrm{var}(K_{100})}{2(17^2)} \approx 0.0433.$$

Using instead the Chernoff bound, $\mathrm{P}(K_{100} \geq 67) \leq g_{\Delta K_{100}}(r)e^{-67r}$. Since K_{100} is binomial, $g_{K_{100}}(r) = (0.5 + 0.5e^r)^{100}$. Since $r > 0$ is arbitrary, we can choose it so to minimize the bound. Specifically, we need

$$0 = \frac{\partial}{\partial r}e^{-67r}(0.5 + 0.5e^r)^{100} = -67e^{-67r}(0.5 + 0.5e^r)^{100} + 100e^{-67r}(0.5 + 0.5e^r)^{99}(0.5e^r).$$

Noticing that $(0.5 + 0.5e^r) > 0$ and $e^{-67r} > 0$, the desired expression becomes $0 = -67(0.5 + 0.5e^r) + 100(0.5e^r)$. Thus, $e^r = 67/33$, or, $r = 0.708$ and the corresponding bound becomes

$$\mathrm{P}(K_{100} \geq 67) \leq 0.002748$$

Both the Chebyshev and the Chernoff bounds are exactly that: bounds. In other words, they provide a guarantee that the desired probability will always be contained within a certain interval, here $[0, 0.0433]$ and $[0, 0.002748]$, respectively. The upper or lower bound need not be a particularly good point estimate. Calculated exactly, $\mathrm{P}(K_{100} \geq 67) = 0.0004369$ to four significant digits, which is two orders of magnitude less than the estimate via the Chebyshev bound and almost an order of magnitude less than the estimate via the Chernoff bound.

The Central Limit Theorem for finite n, on the other hand, is an approximation. In other words, there are no guarantees as to how close the approximate value is to the true answer, nor whether it exceeds it or lies below it, and the quality of the approximation depends vastly on the probabilities of interest (the approximation becomes less and less accurate in the tails of the distribution). However, for many practical situations of interest and including the case at hand, the Central Limit Theorem yields a surprisingly accurate estimate.

Problem 5

a Since the sum of independent random variables is associated with the product of their MGF's, for a class of distributions to be stable, the corresponding class of MGF's needs to be closed under multiplication.

b Stable distributions act as attractors for linear combinations, and thus occur in limiting theorems governing the behavior of sums of IID random variables.

i The Exponential family is not stable. In fact, a sum of two exponentials is an Erlang density.

ii The Erlang family is not stable. Specifically, for two Erlang r.v.'s with means λ_1 and λ_2 and parameter n, the product $(\frac{\lambda_1}{\lambda_1+r})^n(\frac{\lambda_2}{\lambda_2+r})^n$ cannot be expressed as $(\frac{\mu}{\mu+r})^{2n}$ for any $\mu > 0$ and integer $m \geq 1$.

iii The Gaussian family is stable. Specifically, the sum of two independent Gaussians with means μ_1, μ_2 and variances σ_1^2, σ_2^2 is Gaussian with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$ (i.e. $e^{ra+r^2s^2/2}e^{ra+r^2s^2/2} = e^{2ra+r^2s^2}$). Naturally, the Gaussian family shows up in the Central Limit Theorem.

iv The Uniform family is not stable. For example, the density associated with a sum of two IID uniform r.v.'s is a triangle.

v The Binary family is not stable. In particular, the sum of two IID binary r.v.'s is a binomial r.v.

vi The Binomial family is not stable. We see from the product of the MGF's that the sum of two independent binomials with success probabilities p_1 and p_2 on n trials cannot be expressed as a binomial r.v. with some success probability $0 < q \leq 1$ on m trials.

vii The Geometric family is not stable, since $\frac{qe^r}{1-(1-q)e^r}\frac{qe^r}{1-(1-q)e^r} \neq \frac{pe^r}{1-(1-p)e^r}$ for any $0 < p \leq 1$.

viii The Poisson family is stable, since the sum of two independent Poisson r.v.'s with means λ_1 and λ_2 is a Poisson r.v. with mean $\lambda_1 + \lambda_2$. The Poisson distribution is the limit of a sum of a large number of IID Binary distributions, each with a vanishing success probability. This limit is known as the "law of rare events".

ix The Cauchy family is stable as $e^{-a_1|w|}e^{-a_2|w|} = e^{-(a_1+a_2)|w|}$ for $a_1, a_2 > 0$. The Cauchy distribution occurs in a generalization of the Central Limit Theorem according to which the linear combination of a sum of independent random variables whose cumulative distribution function falls off as $1/x$ tends to a Cauchy distribution.

c The Binomial family with $p = 6$ is stable, as the sum of two independent binomials with success probability 0.6 on n_1, n_2 trials, respectively, yields another binomial with success probability 0.6 on $n_1 + n_2$ trials. A similar observation holds for the Erlang family.

d With exception of the Cauchy family, all of the stable families in b) are discrete. Recall that the sum of discrete random variables remains discrete with increasing n, while taking on the *shape* of a Gaussian. It may help to think of a sampled Gaussian. Loosely speaking, we're then simply faced with the fact that there are several stable random variables that sample the Gaussian distribution in the limit. And that's fine. As for the Cauchy distribution, recall that the Central Limit Theorem requires the underlying random variables to have finite mean and variance. Is that the case of the Cauchy distribution?