A1. To choose a full-house hand, we must choose a rank for three cards ($\binom{13}{1}$ possibilities), three cards in that rank ($\binom{4}{3}$), one of the other ranks for two cards ($\binom{12}{1}$), and two cards in that rank ($\binom{4}{2}$). Therefore the probability is

$$\frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}}.$$

A2. If we have four cards forming two pairs, then there are 48 remaining cards in the deck, and 4 of those cards will complete our hand to a full house. Therefore the probability is $4/48 = 1/12$.

B1. [We discussed this problem recently in class, because it came up in the practice problems.] Let G_1 be the number of trials needed for the first success, G_2 the number of additional trials (after the first G_1 trials) needed for the second success, G_3 the number of additional trials (after the first $G_1 + G_2$ trials) needed for the third success, etc. Then $X = G_1 + \cdots + G_r$. Each G_k is a geometric random variable. The G_k are independent, because the geometric distribution is memoryless (in other words, because the underlying Bernoulli trials are independent).

B2. In class we showed that the MGF for each G_k is

$$m_{G_k}(t) = \frac{pe^t}{1 + e^t(p-1)}.$$

Then, because the G_k are independent and X is their sum,

$$m_X(t) = \left(\frac{pe^t}{1 + e^t(p-1)}\right)^r.$$

C. Let D be the event that I have the disease and T the event that I test positive for it. We are told that $P(T|D) = P(T^c|D^c) = 0.95$. Also, $P(D) = 0.01$ is the probability that I have the disease before receiving any test results. We wish to find

$$\begin{aligned} P(D|T) &= \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)} \\ &= \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} \\ &= \frac{95 \cdot 1}{95 \cdot 1 + 5 \cdot 99} \\ &= 95/590 \\ &= 19/118. \end{aligned}$$

[By the way, that's about 16%.]

D. We begin by computing the CDF of Z:

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(X + Y \leq z) \\ &= \iint_R f(x, y)\, dx\, dy, \end{aligned}$$

where R is the set of all points (x, y) such that $x + y \leq z$. [A good solution is accompanied by a diagram showing R and thus explaining the following integration bounds. I omit the diagram in these typed solutions.] The integral equals

$$\int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y)\, dy\, dx.$$

Therefore the density of Z is

$$\begin{aligned} f_Z(z) &= \frac{d}{dz} F_Z(z) \\ &= \frac{d}{dz} \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y)\, dy\, dx \\ &= \int_{-\infty}^{\infty} \frac{d}{dz} \int_{-\infty}^{z-x} f(x, y)\, dy\, dx \\ &= \int_{-\infty}^{\infty} f(x, z - x)\, dx \end{aligned}$$

by the fundamental theorem of calculus.

E. Let $S_n = X_1 + \cdots + X_n$ as usual. We wish to find n such that

$$\begin{aligned} 0.99 &= P\left(\left|\frac{S_n}{n} - \mu\right| \leq \epsilon\right) \\ &= P\left(-\epsilon \leq \frac{S_n}{n} - \mu \leq \epsilon\right) \\ &= P\left(-\frac{\epsilon}{\sigma/\sqrt{n}} \leq Z \leq \frac{\epsilon}{\sigma/\sqrt{n}}\right), \end{aligned}$$

where

$$Z = \frac{\frac{S_n}{n} - \mu}{\sigma/\sqrt{n}}$$

is a standard normal random variable (approximately, for large n, by the central limit theorem). Let F be the CDF of Z. The probability above equals 0.99 when $F(\epsilon\sqrt{n}/\sigma) = 0.995$. Therefore the minimum usable n is

$$n = \left(\frac{\sigma}{\epsilon} F^{-1}(0.995)\right)^2.$$

[In practice, σ would be estimated from the data, and ϵ (and the 0.99) would be chosen according to the biologist's tastes and needs.]

F. Recall from class that the CDF of X is $F(x) = 1 - e^{-\lambda x}$. Setting $u = F(x)$ and solving for x as a function of u gives

$$x = -\frac{1}{\lambda} \log(1 - u).$$

So we choose u uniformly on $[0, 1]$ and compute x from u using that equation.

G1. Well,

$$\begin{aligned}
F_{X|X>s}(x|x > s) &= P(X \leq x|X > s) \\
&= \frac{P(X \leq x, X > s)}{P(X > s)} \\
&= \frac{\int_s^x f(t)\, dt}{1 - F(s)} \\
\Rightarrow \qquad f_{X|X>s}(x|x > s) &= \frac{d}{dx} F_{X|X>s}(x|x > s) \\
&= \frac{\frac{d}{dx} \int_s^x f(t)\, dt}{1 - F(s)} \\
&= \frac{f(x)}{1 - F(s)}
\end{aligned}$$

by the fundamental theorem of calculus. By the way, the domain of $f_{X|X>s}(x|x > s)$ is (s, ∞) rather than $(-\infty, \infty)$. In other words, $f_{X|X>s}(x|x > s) = 0$ for $x \leq s$. This claim should make intuitive sense. You can also verify it by computing

$$\int_s^\infty f_{X|X>s}(x|x > s)\, dx = \frac{1}{1 - F(s)} \int_s^\infty f(x)\, dx = \frac{1}{1 - F(s)} (1 - F(s)) = 1.$$

G2. [Because I made an error while writing this problem, it requires more computation than I intended. Grading was generous.] We are asked to compute

$$\begin{aligned}
E[X|X > 1] &= \int_1^{10} x f_{X|X>1}(x|x > 1)\, dx \\
&= \frac{1}{1 - F(1)} \int_1^{10} \frac{x f(x)}{1 - F(1)}\, dx \\
&= \frac{1}{1 - F(1)} \int_1^{10} x f(x)\, dx \\
&= \frac{1}{1 - F(1)} \int_1^{10} x 0.7 (0.2x - 1)^6\, dx.
\end{aligned}$$

There are two subtasks. First,

$$F(1) = \int_0^1 f(x)\, dx = \int_0^1 0.7(0.2x - 1)^6\, dx = \left[\frac{1}{2}(0.2x - 1)^7\right]_0^1 = -\frac{1}{2} 0.8^7 + \frac{1}{2}.$$

Second, using integration by parts,

$$\begin{aligned}
\int x 0.7(0.2x - 1)^6\, dx &= \frac{1}{2} x (0.2x - 1)^7 - \int \frac{1}{2} (0.2x - 1)^7\, dx \\
&= \frac{1}{2} x (0.2x - 1)^7 - \frac{1}{3.2} (0.2x - 1)^8 + C.
\end{aligned}$$

Therefore

$$\begin{aligned} E[X|X>1] &= \frac{1}{1+\frac{1}{2}0.8^7-\frac{1}{2}}\left[\frac{1}{2}x(0.2x-1)^7-\frac{1}{3.2}(0.2x-1)^8\right]_1^{10} \\ &= \frac{1}{1+0.8^7}\left[x(0.2x-1)^7-\frac{1}{1.6}(0.2x-1)^8\right]_1^{10} \\ &= \frac{10-\frac{1}{1.6}+0.8^7+\frac{1}{1.6}0.8^8}{1+0.8^7}. \end{aligned}$$

[By the way, that's approximately 8.01. So surviving the first year greatly increases the expected life span, from 5 to 8 years. Similar remarks apply to the life spans of humans.]