Stat 110 Homework 10   Mark Grozen-Smith

1. a) We know that the random variable $Y|I_A = I_A(Y|A) + (1 - I_A)(Y|A^c)$ just because of how A and its indicator are related. Taking the expected value of both sides yields $\boxed{E(Y|I_A) = I_A E(Y|A) + (1 - I_A)E(Y|A^c)}$

   b)

$$P(YI_A = y) = \begin{cases} P(yI_A = y|A)P(A) + (1 - P(A)), & \text{if } y=0 \\ P(yI_A = y|A)P(A), & else \end{cases}$$

$$E(YI_A) = \Sigma_1^\infty yP(yI_A = y|A)P(A)$$

Since we are saying given A in the expression, we can ignore the indicator

$$E(YI_A) = \Sigma_1^\infty yP(y = y|A)P(A)$$

Notice this is the expected value of Y given A times P(A)!

$$E(YI_A) = E(Y|A)P(A) \rightarrow E(Y|A) = \frac{E(YI_A)}{P(A)}$$

   c) We just showed that $E(YI_A) = E(Y|A)P(A)$, and this was for I_A indicating that A did happen. Let's do the same thing for I_A^* which indicates that A did NOT happen. Then, $E(YI_A^*) = E(Y|A^c)P(A^c)$ This means that we can create the entire domain of Y with $Y = YI_A + YI_A^*$. If we take the expectation of both sides, we get

$$\boxed{E(Y) = E(YI_A) + E(YI_A^*) = E(Y|A)P(A) + E(Y|A^c)P(A^c)}$$

2. a) The time until the first person shows up, T_1, can be modeled as a beta since the guests arrive uniformly. $T_1 \ Beta(1, N)$. This is useful because then the expected time until the first guest arrives, using how many people show up in total, N, is as shown below.

$$E(T_1) = E(E(T_1|N)) = E(\frac{1}{1+N}) = \Sigma_{n=1}^\infty \frac{1}{1+N}p(N = n|N \geq 1)$$

Since N is Poisson, we know $p(N = n|N \geq 1) = \frac{e^{-\lambda}\lambda^n}{n!(1-e^{-\lambda})}$. With a little massaging, this gives

$$E(T_1) = \frac{e^{-\lambda}}{\lambda(1 - e^{-\lambda})}\Sigma_2^\infty \frac{\lambda^n}{n!}$$

With a Taylor series substitution for e^x

$$E(T_1) = \frac{e^{-\lambda}}{\lambda(1 - e^{-\lambda})}(e^{-\lambda} - 1 - \lambda)$$

For $\lambda = 20$ over the 240 minutes interval we have here, this turns into $\boxed{\text{12 minutes after 8pm.}}$

b) By symmetry, this question is exactly the same as part (a), but instead of estimating how many minutes after 8 pm the first person arrives, we'll be predicting how many minutes before 12 the last person arrives. This leaves us with $\boxed{\text{12 minutes before 12.}}$

3. a) I hope you're feeling originally sinful because Adam and Eve are back to play.

$$E(N) = E(E(N|\lambda)) = E(\lambda) = \boxed{1}$$

$$Var(N) = Var(E(N|\lambda)) + E(Var(N|\lambda)) = Var(\lambda) + E(\lambda) = 1 + 1 = \boxed{2}$$

b) Let's define a variable $T|N$ that is the total amount spent given the number of claims. This is a useful r.v. to create because, if the number of claims is fixed, then we know the average amount spent in total $(E(T|D))$ is just the average price of a claim times the number of claims, $N * E(D)$ Then, Adam's Law helps us out again (as does the book where I looked up the mean of a log-normal distribution, and the fact that N and D are independent).

$$E(T) = E(E(T|N)) = E(N * E(D)) = E(N)E(D) = \boxed{e^{\mu + \frac{\sigma^2}{2}}}$$

Again, Eve is killin it when we try to find the variance here.

$$Var(T) = Var(E(T|N)) + E(Var(T|N)) = Var(N*E(D)) + E(N)*(e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$$

$$\boxed{Var(T) = 2(e^{\mu + \frac{\sigma^2}{2}}) + (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}}$$

c) We know the PMF of N given λ is $P(N = n|\lambda) = \frac{e^{-\lambda}\lambda^n}{n!}$ and then making this conditional distribution into a joint distribution, $P(N = n|\lambda = l) = \frac{e^{-\lambda}\lambda^n}{n!} * e^{-l}$ In order to get the distribution of just N, we need to integrate out the λ.

$$P(N = n) = \int_0^\infty P(N = n|\lambda = l)dl = \int_0^\infty \frac{e^{-\lambda}\lambda^n}{n!} * e^{-l}dl$$

2

$$P(N = n) = \frac{1}{2^n n!} \int_0^\infty e^{-2l}(2l)^n dl$$

It's clever, but we can use the fact that $\Gamma(n) = \int_0^\infty t^{n-1} e^{-t} dt$ to our advantage here. Setting t = 2d we get

$$P(N = n) = \frac{1}{2^n n!} \int_0^\infty e^{-t} t^n \frac{1}{2} dt = \frac{\Gamma(n+1)}{2^n n!}$$

$$\boxed{P(N = n) = \frac{1}{2^{n+1}} = Geom(\frac{1}{2})}$$

d) By simple Bayes we know that $P(\lambda = l|N = n) = \frac{P(N=n|\lambda=l)P(\lambda=l)}{P(N=n)}$

$$P(\lambda = l|N = n) = \frac{\frac{e^{-l}l^n}{n!}e^{-l}}{\frac{1}{2^{n+1}}} = \boxed{\frac{2^{n+1}e^{-2l}l^n}{n!}}$$

This we can extract to see that it it just Gamma(n+1,2)

4. a) The expected value of $\bar{X}^*|X_1, \ldots, X_n$ is simply the same as the expected value for all the X variables, $\bar{X}$. This is because the expected value of each the X^* variables, given our set of X's, is just the mean of the X variables. Each X variable has an equal probability of being chosen for each X^*, so the expected value of each X^* is just the average of the Xs. Further, since the X^*s are all independent and contribute eqaully to the mean of them, the expected value of $\bar{X}^*$ is just the simple mean of the expected values, which comes out to $\boxed{E(\bar{X}^*|X_1, \ldots, X_n)) = \bar{X}}$.

For the variance, we know that $Var(\bar{X}^*|X_1, \ldots, X_n) = Var(\frac{1}{n}\Sigma_{i=1}^n X_i^*|X_1, \ldots, X_n)$. Since these are independent, we can just sum up their variances, and since they are identically distributed, we know that means just multiplying by n. Doing that and taking the constant out of the variance expression, we get

$$\frac{n}{n^2} Var(X_i^*|X_1, \ldots, X_n)$$

Which is just the variance of the set of Xs divided by n which leaves us with

$$\boxed{Var(\bar{X}^*|X_1, \ldots, X_n) = \frac{\Sigma_{i=1}^n (X_i - \bar{X})^2}{n^2}}$$

b) By Adam's Law, $E(\bar{X}^* = E(E(\bar{X}^*|X_1, \ldots, X_n))$ which is just $E(\bar{X})$. We know this is $\boxed{\mu}$ because the mean of the distribution of X is μ so over many samples we know the mean of this set will be the same.

3

Again that darn variance is a bit more an issue, but that's ok. Eve knows what's up.

$$Var(\bar{X}^*) = Var(E(\bar{X}^*|X_1, \ldots, X_n)) + E(Var(\bar{X}^*|X_1, \ldots, X_n))$$

$$Var(\bar{X}^*) = Var(\bar{X}) + E(\frac{\Sigma_{i=1}^n (X_i - \bar{X})^2}{n^2})$$

We know the first expression here is just $\frac{Var(X_i)}{n}$ since, again, this average is linear over independent variables so we can pull out the $\frac{1}{n}$ constant and multiply by n on the variance of the X distribution. The second expression is just the definitely of the expected value of the sample variance leaving us with

$$\boxed{Var(\bar{X}^*) = \frac{\sigma^2}{n} + \frac{\sigma^2(n-1)}{n^2}}$$

c) These are not the same distributions. The most obvious was to see this is that the tails of the X^* distribution are heavier than those of the X distribution. This is because of rare events in the X distribution and how often they happen in the X^* distribution. Let's say X has a certain value once in every hundred pulls or so ($P(X = a) = \frac{1}{100}$). This is rare and will only happen every now and then, and when it does happen, it will only happen once since we draw a number for X and then draw a completely independent number. However, when this does happen in X, the X^* distribution has a high probability of hitting that number multiple times, which gives the X^* distribution much thicker tails, a much higher variance, and an overall different distribution.

5. a) Our optimal strategy is dependent on p. If $p > \frac{1}{2}$ then we want to always guess 1. Otherwise, we should always guess 0. With this strategy, we have a probability p of guessing the nth bit correct.

b) By the law of large numbers, the proportion of 1s and 0s will eventually hit the actual probability of getting a 1 or 0. This is the basis of the frequentist approach to statistics, I believe. Thus, we will eventually be estimating with the actual p just like in part a and we will, again, have a probability p of guessing the nth bit correct.

c) Of course, if the demon can change the probability at any moment, that would make things difficult for me. Just have p=0, then once I decide to guess 0, switch p to equal 1. This is essentially the same as keeping my estimated probability equal to one half. Having p = 0.5 means that I have

no information about the bits' probability. This would make it very hard for me to have any confidence in my guesses.

6. a) We know that any two Poisson distributions can be combined into one by just adding their lambda values such that something like $Pois(\lambda_1) + Pois(\lambda_2) = Pois(\lambda_1 + \lambda_2)$. The opposite direction also applies, which is useful here. Now we can say that $Pois(n) = n * Pois(1)$ which is just a sum of i.i.d. r.v.s all of which are $Pois(1)$. By the CLT, we know that we will get a normal out of this summation.

b) At the limit of n as it approaches infinity, we know that the $Pois(n)$ is a normal distribution which means the means and variances must be the same. This gives

$$E(X_n) = n = \mu$$

$$Var(X_n) = n = \sigma^2$$

From here, we take the MGF of the standardization of this distribution.

$$S = \frac{X_n - \mu}{\sigma} = \frac{X_n - n}{\sqrt{n}}$$

MGF of $S = E(e^{t\frac{X_n - n}{\sqrt{n}}}) = e^{-\sqrt{n}t} E(e^{\frac{tX_n}{\sqrt{n}}})$

Since we also know that X_n is a Poisson r.v., we know it's MGF is

$$E(e^{tX_n}) = e^{n(e^t - 1)}$$

After making that substitution we get

$$e^{-\sqrt{n}t} * e^{n(e^{\frac{t}{\sqrt{n}}} - 1)}$$

$$e^{-\sqrt{n}t - n + ne^{\frac{t}{\sqrt{n}}}}$$

From here, we want to get rid of that exponential in the exponential. To do this, we can use Taylor's theorem with remainder that $e^x = 1 + x + \frac{x^2}{2} + O(x^3)$. In this example, $e^{\frac{t}{\sqrt{n}}} = 1 + \frac{t}{\sqrt{n}} + \frac{t^2}{2n} + O(\frac{t}{\sqrt{n}})$ which we can plug in to our expressiong for the MGF of S to yield

$$e^{-\sqrt{n}t - n + n(1 + \frac{t}{\sqrt{n}} + \frac{t^2}{2n} + O(\frac{t}{\sqrt{n}}))} = e^{-\sqrt{n}t - n + n + t\sqrt{n} + \frac{t^2}{2} + nO(\frac{t}{\sqrt{n}})}$$

Since that big-O function is 0 for sufficiently large n (because a large n leads to a small argument to that big-O function), we can drop that term as we

take the limit of n to infinity. After dropping that term and canceling out the rest, this leaves

$$e^{\frac{t^2}{2}}$$

Which IS the MGF for the normal and thus we have shown, without invoking the great CLT, that $Pois(n)$ is normal as n approaches infinity.