24|12|21

Agenda: - Assignment 2 Q 1, 4, 7.
- Ross Ch 5, Ex 2(c), 4(c), 4(i), 5(f), 7(a).
- A brief note on "rigorous" prob theory.

Assignment Q...

P1) This follows from the definition of the Poisson(λ) random variable. That is,

$$\frac{\mathbb{P}(X=k)}{\mathbb{P}(X=k-1)} = \frac{e^{-\lambda}\lambda^k/k!}{e^{-\lambda}\lambda^{k-1}/(k-1)!} = \lambda/k$$

Now if [illegible] integers then $\mathbb{P}(X=k) = \mathbb{P}(X=k-1)$ is the maximum [illegible]. Keep $\lambda > k \Rightarrow \mathbb{P}(X=k)$ is the maximum.

For the expectation,

$$\mathbb{E}[f(X)] = \sum_{k=1}^{\infty} \ldots = \sum_{k=1}^{\infty} k f(k) \mathbb{P}(X=k) = \ldots = \lambda \sum f(k+1)\mathbb{P}(X=k) = \mathbb{E}[\lambda f(X+1)]$$

Key point: The "mode" of the Poi(λ) rv is very close to its mean λ. This means that Poi(λ) concentrates heavily around its mean, so you expect mean & variance to be close [illegible]. Indeed, $\mathbb{E} = \mathrm{Var} = \lambda$, so it's binomial-like concentration!

(Q4) Consider the situation $f_X(n) = Cn^{-\alpha}$. Then, for this to be the pmf of a $\mathbb{N}$-valued random variable, we require that $\sum_{n\geq 1} f_X(n) = 1$, which becomes $C\sum n^{-\alpha} = 1$.

By the integral test, $\sum n^{-\alpha}$ converges to some $K \in \mathbb{R}$. Then $C = 1/K$.

Now, note that $\mathbb{E}[X] = \sum C n \cdot n^{-\alpha} = C\sum n^{1-\alpha}$. Noting that $-\alpha+1 < -1$, we get that $\sum n^{1-\alpha}$ converges to a finite number. So $\mathbb{E}[X]$ is well-defined.

The point: Markov's inequality tells you that if $\mathbb{E}[X]$ [illegible] then $\mathbb{P}(X>t) \leq \mathbb{E}[X]/t$. In other words, if $\mathbb{E}[X]$ is finite, then $\mathbb{P}(X>t)$ decays like $1/t$. Here, we have the converse: if $\mathbb{P}(X>t)$ decays like [illegible], then $\mathbb{E}[X]$ exists. Indeed for $f_X(n)$, $\mathbb{P}(X>n) \approx$ [illegible], so this illustrates the sense of the Markov inequality, in some sense.

(P7) First caveat: Using Fubini-Tonelli theorem here is a **FORMALITY**: it's a one-line obvious justification: you can always exchange summation when all terms are non-negative. So there is no need to worry here.

The idea is: indicators.

a)
$$\sum_{k=1}^{\infty} (a_1+\cdots+a_k)\mathbb{P}(X=k) = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} a_j 1_{\{k\geq j\}}\mathbb{P}(X=k) = \ldots = \sum_{j} a_j \mathbb{P}(X\geq j)$$

For (b), use $a_i = 1$ ∀ i above. For (c), use $a_i = 2i$ above for all i, along with the fact that

$$\sum_{j=1}^{\infty}(a_1+\cdots+a_j)\mathbb{P}(X=j) = \mathbb{E}(f(X))$$
where $f(j) = a_1+\cdots+a_j$

Ross

2(c) Points to take away:
(i) Usage of the change of variable formula.
(ii) The logic that is used to find the random variable of interest: identifying it as a function of the underlying randomness.
(iii) Interesting question: is it possible to prove that $\mathbb{E}[L_p(U)] \geq \mathbb{E}[L_q(U)]$ $\forall\, |q-1/2| \geq |p-1/2|$? WITHOUT looking at the distribution of L_p and L_q?

4(c) Key takeaways:
(i) Exam marks can be modeled using a normal rv: to be more precise, the reason is that if independent questions are taken as Bernoullis of various unknown ps, then their sum can be approximated by a normal random variable, suitably normalized.
(ii) Once you assume your data is normal, then there is a reasonable scheme of grading: this can be thought of as "discretizing" the normal, since the grades form a discrete set, but are a function of the normal.

4(i) Key takeaways:
(i) **Comprehension**: This question aims to test your understanding of the scenario and see "which" random variable is applicable, and how. Once you see it is the binomial, you are home. The fact that the diet has no effect on the chol. level tells you that each person's chol. level is likely to go any way (up or down) during the experiment, i.e. Ber (1/2) for each of 100 independent people.
(ii) Using the normal approximation to the Binomial:

$$\left[\frac{\mathrm{Bin}(n,p) - np}{\sqrt{np(1-p)}} \approx N(0,1) \text{ [in distribution]}\right]$$

Ex5(f): Key takeaways:
→ Understanding the hazard rate concept: The best explanation is that:

$$\mathbb{P}[X < t+\Delta t \mid X > t] \approx \lambda_X(t)\Delta t \quad [\forall \Delta t \approx 0]$$

→ The Hazard Rate inversion formula:

$$F(t) = 1 - \exp\left\{-\int_0^t \lambda_X(s)\,ds\right\}$$

[Remember that this works only for non-neg. random variables]

This extends to the more general:

$$\mathbb{P}[X>B \mid X>A] = \exp\left\{-\int_A^B \lambda_X(t)\,dt\right\}$$

→ Interpreting the hazard rate when it appears in a question. The key word is rate here.

→ When you consider random variables that model "lifetimes" in general, it may be required to compare random variables that represent such events. One of these is the notion of stochastic dominance. Another, is the hazard rate order: so we can compare random variables reasonably using this quantity.

7(a) Key takeaways:
(i) If f is an increasing 1-1 function, then employing

$$\mathbb{P}(f(X) \geq y) = \mathbb{P}(X \geq f^{-1}(y))$$

yields, instantly, the PDF \ CDF of $F(X)$ in terms of ~~that~~ for X.
(ii) If f is strictly decreasing, the above can be repeated in the opposite direction. The short of it is the importance of the above technique.

On rigorous probability theory

It is fair to wonder how the theory of sample spaces and random variables are expanded so as to fit infinite spaces and continuous random variables.

The answer, is to redefine a prob. space as follows:
- A sample space Ω of outcomes.
- A suitable set of subsets of Ω (or events), that we wish to study across our experiment. Call this σ.
- A way to study these, i.e. a "measure" $\mu: \sigma \to [0,1]$ that reasonably assigns probabilities to each such event.

Given these three, you put them together to get a probability space.

The only condition that σ must satisfy is some consistency conditions. Heuristically, σ should be closed under complement & unions, since the ability to study an event implies we can study its complement. Such sets are called σ-algebras.

What happens in the discrete case is that every subset can be studied. Typically, on $\mathbb{R}$, and $[0,1]$ and other non-discrete sample spaces, you usually cannot study every subset if you wish to measure them in a certain manner. Thus, for some measures, you need to choose suitable σ-algebras for these measures.

For $\mathbb{R}$, or $[0,1]$, the respective σ-algebra used is called the Borel σ-algebra. This is the σ-algebra generated by all intervals of the form $\{[a,b] : a \leq b\}$. [generated means to take every set that can be generated from intervals using only unions and complements]

Remarkably enough, it can be shown that any measure, if defined on a generating set, extends to the entire σ-algebra uniquely. Using this, if we say:

$$\Omega = \mathbb{R},\ \sigma = \mathrm{Borel}[\mathbb{R}],\ \mu([a,b]) = \int_a^b e^{-x^2/2}\frac{1}{\sqrt{2\pi}}\,dx$$

Then μ is the prob. measure corresponding to a $N(0,1)$ r.v. since if $X \sim N(0,1)$ then $\mu[a,b] = \mathbb{P}[X \in [a,b]]$.

$$\Omega = [0,1],\ \sigma = \mathrm{Borel}(\mathbb{R}),\ \mu([a,b]) = b-a \Rightarrow \mu \sim U[0,1]$$

$$\Omega = [n]^{\cup\{0\}},\ \sigma = \mathcal{P}([n]^{\cup\{0\}}) \left(\text{generated by } \{\{k\} : k \in [n]^{\cup\{0\}}\}\right),$$
$$\mu(\{k\}) = p^k(1-p)^{n-k}\binom{n}{k}$$

Then $\mu \sim \mathrm{Bin}(n,p)$

Thus, it is possible now to define r.v.s. continuously. More on this in Klenke, Chapter 1.