Math 154: Probability Theory, HW 2

DUE FEB 6, 2024 BY 9AM

Remember, if you are stuck, take a look at the lemmas/theorems/examples from class, and see if anything looks familiar.

1. SOME PRACTICE

1.1. **Poisson and binomial distributions show up everywhere.** Let X and Y be independent Poisson random variables with parameters λ and μ, respectively.

(1) By computing the pmf of $X + Y$, show that $X + Y$ is a Poisson random variable with parameter $\lambda + \mu$

(2) By computing $\mathbb{P}(X = k|X+Y = n)$, show that $\mathbb{P}(X = k|X+Y = n) = p(k)$, where $p(k)$ is the pmf for a Binomial distribution (with parameters that you must compute).

Solution:

(1) Consider the sum $X + Y$.

$$\begin{aligned}
\mathbb{P}(X + Y = k) &= \sum_{i=0}^{k} \mathbb{P}(X + Y = k|X = i)\mathbb{P}(X = i) \\
&= \sum_{i=0}^{k} \mathbb{P}(Y = k - i)\mathbb{P}(X = i) \\
&= \sum_{i=0}^{k} e^{-\mu} \frac{\mu^{k-i}}{(k-i)!} \cdot e^{-\lambda} \frac{\lambda^i}{i!} \\
&= \frac{e^{-\mu-\lambda}}{k!} \sum_{i=0}^{k} \frac{\mu^{k-i}}{(k-i)!} \cdot \frac{\lambda^i}{i!} \\
&= \frac{e^{-\mu-\lambda}}{k!} \sum_{i=0}^{k} \frac{\mu^{k-i}}{(k-i)!} \cdot \frac{\lambda^i}{i!} \\
&= \frac{e^{-\mu-\lambda}}{k!} \sum_{i=0}^{k} \binom{k}{i} \cdot \lambda^i \mu^{k-i} \\
&= \frac{(\mu+\lambda)^k e^{-\mu-\lambda}}{k!}
\end{aligned}$$

which is precisely the Poisson PMF with param $\lambda + \mu$.

(2)

$$\begin{aligned}\mathbb{P}(X=k|X+Y=n) &= \frac{\mathbb{P}(X+Y=n|X=k)\mathbb{P}(X=k)}{\mathbb{P}(X+Y=n)} \\ &= \frac{\mathbb{P}(Y=n-k)\mathbb{P}(X=k)}{\mathbb{P}(X+Y=n)} \\ &= \frac{e^{-\mu}\frac{\mu^{n-k}}{(n-k)!}\cdot e^{-\lambda}\frac{\lambda^k}{k!}}{\frac{(\mu+\lambda)^n e^{-\mu-\lambda}}{n!}} \\ &= \frac{n!}{k!(n-k)!}\cdot\left(\frac{\mu}{\mu+\lambda}\right)^{n-k}\left(\frac{\lambda}{\mu+\lambda}\right)^k,\end{aligned}$$

which is the binomial pmf with parameters n and $\frac{\lambda}{\mu+\lambda}$.

1.2. **What?** Suppose X is a geometric random variable. Show that $\mathbb{P}(X=n+k|X>n)=\mathbb{P}(X=k)$ for any integers $k,n\geqslant 1$. (This is called the "memorylessness" property.)
Solution: By the definition of conditional probability, we have that, for $n,k\geqslant 1$,

$$\begin{aligned}\mathbb{P}(X=n+k|X>k) &= \frac{\mathbb{P}(X=n+k,X>n)}{\mathbb{P}(X>k)} = \frac{\mathbb{P}(X=n+k)}{\mathbb{P}(X>n)} \\ &= \frac{(1-p)^{n+k}}{\sum_{\ell=n+1}^{\infty}(1-p)^{\ell}} = \frac{(1-p)^{n+k}}{(1-p)^{n+1}p^{-1}} = (1-p)^{k-1}p,\end{aligned}$$

and memorylessness is proven.

1.3. **For some reason, probabilists like urns.** An urn contains N balls, b of which are blue and $r=N-b$ of which are red. Let us randomly take n of the N balls (without replacement). If R is the number of red balls drawn, explain briefly why

$$\mathbb{P}(R=k)=\frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}.$$

This is the *hypergeometric distribution*. Now, take the limit $b,N,r\to\infty$ and suppose $\frac{r}{N}\to p$ (and thus $\frac{b}{N}\to 1-p$). (In words, keep a constant fraction of blue and red balls.) Compute the limit of $\mathbb{P}(R=k)$ as $N\to\infty$, i.e. confirm that

$$\mathbb{P}(R=k)\to\binom{n}{k}p^k(1-p)^{n-k}. \tag{1.1}$$

(This is saying that in the limit of infinitely many balls, sampling with and without replacement is the same, as long as the number of samples n and the proportion of colors is fixed.) (*Hint*: it may help to use $\frac{\binom{X}{y}}{\frac{1}{y!}X^y}\to 1$ as $X\to\infty$ and y is fixed, i.e. not large.)
Solution:

(1) $\mathbb{P}(R=k)$ can be determined by taking the total number of possible combinations of k red balls and $n-k$ blue balls and dividing it by the total number of possible combinations. There are a total of $\binom{r}{k}$ possible combinations of red balls that can be

chosen and $\binom{N-r}{n-k}$ possible combinations of blue balls that can be chosen. The total number of combinations is $\binom{N}{n}$, giving us the desired equality.

(2) First, we prove the hint. Expanding $\binom{X}{y}$ gives us:

$$\binom{X}{y} = \frac{X!}{(X-y)! \cdot y!} = \frac{1}{y!} \cdot \frac{X!}{(X-y)!} = \frac{1}{y!} \cdot X \cdot (X-1) \cdot \ldots \cdot (X-y+1).$$

Now, as X grows larger, any constant subtracted from X will just be X, giving us that as X approaches ∞, $\binom{X}{y} = \frac{1}{y!} \cdot X^y$.

Now, using this hint, we begin by adjusting our original equality for $\mathbb{P}(R = k)$. Because r approaches ∞, we have that $\binom{r}{k} = \frac{1}{k!} \cdot r^k$. Applying the same logic to the rest of the equality gives us that

$$\mathbb{P}(R = k) = \frac{\frac{1}{k!} r^k \cdot \frac{1}{(n-k)!} \cdot (N-r)^{n-k}}{\frac{1}{n!} N^n}.$$

Now, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, so we can simplify to have

$$\mathbb{P}(R = k) = \binom{n}{k} \cdot \frac{r^k \cdot (N-r)^{n-k}}{N^n}.$$

Now, given that $p = \frac{r}{N}$, we then have that

$$\mathbb{P}(R = k) = \binom{n}{k} \cdot p^k \cdot \left(\frac{N-r}{N}\right)^{n-k} = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

2. Some lemmas (and another urn)

2.1. **The "layer-cake formula" (and an application).**

(1) Suppose X is a discrete random variable that takes values in the non-negative integers. Show that $\mathbb{E}(X) = \sum_{n=0}^{\infty} \mathbb{P}(X > n)$.

(2) An urn contains b blue and r red balls. Balls are removed from the urn at random one-by-one. Compute the expected number of turns that we must take until the first red balls is drawn. (You should get $\frac{b+r+1}{r+1}$, but show your work.)

Solution:

(1) Note that (smoothing over some of the formalities), we can say

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x=0}^{\infty}\sum_{k=0}^{x}\mathbb{P}(X=x)\\ &= \sum_{x=0}^{\infty}\sum_{k=0}^{\infty}I(x>k)\mathbb{P}(X=x)\\ &= \sum_{k=0}^{\infty}\sum_{x=0}^{\infty}I(x>k)\mathbb{P}(X=x)\\ &= \sum_{k=0}^{\infty}\mathbb{P}(X>k)\end{aligned}$$

(2) Enumerate the b blue balls with indexes 1 through b, defining an indicator I_i equal to 1 if blue ball i is drawn before any red ball. Thus, we have

$$X-1=\sum_{i=1}^{b}I_i$$

For any given blue ball i, considering it alongside all the r red balls, the probability that the blue ball happens before all the reds is $\frac{1}{r+1}$ (since each ordering of $r+1$ balls is equally likely). Thus,

$$\mathbb{E}[X]-1=\sum_{i=1}^{b}\mathbb{E}[I_i]\Rightarrow\mathbb{E}[X]=1+b\cdot\frac{1}{r+1}=\frac{b+r+1}{r+1}.$$

KY: In terms of using the layer cake formula, what is happening here is that $\mathbb{P}[X>i]=\frac{1}{r+1}$ for all $i=1,\ldots,b$ and 0 for $i>b$. So $\mathbb{E}[X]=\mathbb{P}[X>0]+\frac{b}{r+1}$, but $\mathbb{P}[X>0]=1$, so $\mathbb{E}[X]=1+\frac{b}{r+1}=\frac{b+r+1}{r}$.

2.2. **Maximum disorder.** Let $X_1,\ldots,X_n$ be independent Bernoulli random variables with parameters $p_1,\ldots,p_n\in[0,1]$, respectively. Define $Y=X_1+\ldots X_n$.

(1) Show that $\mathbb{E}(Y)=\sum_{k=1}^{n}p_k$ and $\mathrm{Var}(Y)=\sum_{k=1}^{n}p_k(1-p_k)$.
(2) Suppose we fix the value of $\mathbb{E}(Y)$ (to be, say, E). Show that the choice of $p_1,\ldots,p_n$ which maximizes $\mathrm{Var}(Y)$ satisfies $p_1=\ldots=p_n$. (This part has nothing random in it. You can do it by Lagrange multipliers or by plugging in $p_n=E-(p_1+\ldots+p_{n-1})$ into the variance formula and maximizing over $n-1$ variables without any constraints by using calculus.)

Solution:

(1) Using linearity of expectation, we have that

$$\mathbb{E}[Y]=\mathbb{E}[X_1]+\mathbb{E}[X_2]+...+\mathbb{E}[X_n]=p_1+p_2+...+p_n.$$

Because $X_1,...,X_n$ are independent variables, we have that

$$\mathrm{Var}(Y)=\mathrm{Var}(X_1)+\mathrm{Var}(X_2)+...+\mathrm{Var}(X_n)=p_1(1-p_1)+...+p_n(1-p_n).$$

(2) Suppose $\mathbb{E}[Y] = \sum_{k=1}^n p_k = E$. Then $p_n = E - \sum_{k=1}^{n-1} p_k$. This means the variance is equal to

$$\sum_{k=1}^{n-1} p_k(1-p_k) + \left(E - \sum_{j=1}^{n-1} p_j\right)\left[1 - E + \sum_{j=1}^{n-1} p_j\right].$$

Let's take the p_k derivative. We get

$$1 - 2p_k - 1 + 2E - 2\sum_{j=1}^{n-1} p_j.$$

We want all of these to be 0, which means that

$$p_k = E - \sum_{j=1}^{n-1} p_j$$

for all $k = 1, \ldots, n-1$. But this is the value of p_n too, so we are done. (Technically, one has to check that the maximum of $\text{Var}(Y)$ cannot occur on the boundary, i.e. when $p_k = 0$ or $p_k = 1$ for some k. But in this case, the variance is 0, whereas clearly the variance does not have to be 0 if the expectation is fixed, unless $E = n$ or $E = 0$, but this is case is trivial. But, maybe this isn't necessary for grading purposes.)

2.3. **An old friend, the covariance matrix.** Let $X_1, \ldots, X_n$ be (possibly dependent) random variables. Define the matrix $\text{Cov}(\mathbf{X})$ as an $n \times n$ matrix with entries $\text{Cov}(\mathbf{X})_{ij} = \text{Cov}(X_i, X_j)$. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be the vector with entries $X_1, \ldots, X_n$.

(1) Show that for any $\mathbf{v} = (v_1, \ldots, v_n)$, we have $\mathbf{v}\text{Cov}(\mathbf{X})\mathbf{v}^T = \text{Var}(\mathbf{v} \cdot \mathbf{X}) \geqslant 0$.
(2) Show that $\text{Cov}(\mathbf{X})$ is invertible if and only if the following is satisfied:
   - If $\mathbf{v} \cdot \mathbf{X}$ is a constant random variables, then $\mathbf{v}$ is the zero vector.

   (*Hint*: what condition on the null-space is equivalent to a square matrix being invertible?)

Solution:

(1) By bilinearity of the covariance,

$$\text{Var}(\mathbf{v}\cdot\mathbf{X}) = \text{Var}\left(\sum_{i=1}^n v_i X_i\right) = \text{Cov}\left(\sum_{i=1}^n v_i X_i, \sum_{j=1}^n v_j X_j\right) = \sum_{i=1}^n \sum_{j=1}^n v_i v_j \text{Cov}(X_i, X_j).$$

Also, the ith entry in $\text{Cov}(\mathbf{X})\mathbf{v}^T$ is

$$\sum_{j=1}^n v_j \text{Cov}(X_i, X_j),$$

so

$$\mathbf{v}\text{Cov}(\mathbf{X})\mathbf{v}^T = \sum_{i=1}^n v_i \sum_{j=1}^n v_j \text{Cov}(X_i, X_j) = \sum_{i=1}^n \sum_{j=1}^n v_i v_j \text{Cov}(X_i, X_j).$$

Hence $\mathbf{v}\text{Cov}(\mathbf{X})\mathbf{v}^T = \text{Var}(\mathbf{v} \cdot \mathbf{X}) \geq 0$, because the variance is always nonnegative.

(2) First assume that $\text{Cov}(\mathbf{X})$ is invertible, and let $\mathbf{v} = (v_1, \ldots, v_n)$ be a vector such that $\mathbf{v} \cdot \mathbf{X} = c$ is a constant. Then, for any i, by linearity of expectation,

$\text{Cov}(X_i, \mathbf{v} \cdot \mathbf{X}) = \text{Cov}(X_i, c) = \mathbb{E}(X_i c) - \mathbb{E}(X_i)\mathbb{E}(c) = c\mathbb{E}(X_i) - c\mathbb{E}(X_i) = 0.$

Therefore by bilinearity of the covariance,

$$0 = \text{Cov}(X_i, \mathbf{v} \cdot \mathbf{X}) = \text{Cov}\left(X_i, \sum_{j=1}^{n} v_j X_j\right) = \sum_{j=1}^{n} v_j \text{Cov}(X_i, X_j).$$

But this is precisely the ith entry in the column vector $\text{Cov}(\mathbf{X})\mathbf{v}^T$, so we deduce that $\text{Cov}(\mathbf{X})\mathbf{v}^T$ is the zero vector. However, $\text{Cov}(\mathbf{X})$ was assumed to be invertible, so it must be the case that $\mathbf{v}^T$ is the zero vector. Hence $\mathbf{v}$ is the zero vector.

Conversely, assume that if $\mathbf{v}$ is a vector satisfying $\mathbf{v} \cdot \mathbf{X} = c$ for some constant c, then it is the zero vector (•). To show that $\text{Cov}(\mathbf{X})$ is invertible, it is equivalent to show that if $\text{Cov}(\mathbf{X})\mathbf{v}^T$ is the zero vector, then $\mathbf{v}^T$ is the zero vector. So, let $\mathbf{v}^T$ be a vector such that every entry in $\text{Cov}(\mathbf{X})\mathbf{v}^T$ is 0. From earlier, we know that the ith entry in $\text{Cov}(\mathbf{X})\mathbf{v}^T$ is

$$\sum_{j=1}^{n} v_j \text{Cov}(X_i, X_j).$$

But, this can be rearranged:

$$\sum_{j=1}^{n} v_j \text{Cov}(X_i, X_j) = \text{Cov}\left(X_i, \sum_{j=1}^{n} v_j X_j\right) = \text{Cov}(X_i, \mathbf{v} \cdot \mathbf{X}).$$

Therefore $\text{Cov}(X_i, \mathbf{v} \cdot \mathbf{X}) = 0$ for all i. It follows that

$$\begin{aligned}
\text{Var}(\mathbf{v} \cdot \mathbf{X}) &= \text{Var}\left(\sum_{i=1}^{n} v_i X_i\right) \\
&= \text{Cov}\left(\sum_{i=1}^{n} v_i X_i, \sum_{j=1}^{n} v_j X_j\right) \\
&= \text{Cov}\left(\sum_{i=1}^{n} v_i X_i, \mathbf{v} \cdot \mathbf{X}\right) \\
&= \sum_{i=1}^{n} v_i \text{Cov}(X_i, \mathbf{v} \cdot \mathbf{X}) \\
&= \sum_{i=1}^{n} v_i \cdot 0 \\
&= 0.
\end{aligned}$$

So, $\mathbf{v} \cdot \mathbf{X}$ is a constant. By (•), we get that $\mathbf{v}$ is the zero vector. Thus $\mathbf{v}^T$ is the zero vector, and so we've shown that $\text{Cov}(\mathbf{X})$ is invertible.