HOMEWORK 7

PSTAT 120A: Summer 2022

Due: 11:00am on Thursday, July 21

Instructor: Ethan P. Marzban

Instructions:

- Please submit your work to Gradescope by no later than the due date posted above.
- Be sure to show your work; correct answers with no supporting work will not be awarded full points.
- 2 randomly selected questions/parts will be graded, but you must still turn in your work for all problems in order to be eligible to earn full credit.

1. A few lectures ago, we encountered the χ^2 distribution. In this problem, we shall investigate this distribution further.

   (a) If $T \sim \chi^2$, write down $f_T(t)$, the p.d.f. of T. (Yes, we did prove this in lecture! You don't need to re-derive the result; just write it down.)

   Solution: From the lecture on Transformations,

   $$f_T(t) = \frac{1}{y\sqrt{2\pi}} e^{-\frac{1}{2}y} \cdot \mathbb{1}_{\{y \geq 0\}}$$

   (b) Identify the χ^2 distribution as a special case of one of our familiar distributions.

   Solution: We need only to focus on the variable part of the p.d.f: note that

   $$\frac{1}{\sqrt{y}} e^{-\frac{1}{2}y} = y^{-1/2} e^{-\frac{1}{2}y} = y^{1/2-1} e^{-1/2}$$

   which allows us to recognize $Y \sim \text{Gamma}(1/2,\ 1/2)$. By the way, we didn't explicitly show this but

   $$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

   meaning we can write the full density as

   $$f_Y(y) = \frac{1}{\sqrt{\pi}} \cdot \frac{1}{\sqrt{2}} \cdot y^{1/2-1} e^{-\frac{1}{2}y} \cdot \mathbb{1}_{\{y \geq 0\}} = \frac{\left(\frac{1}{2}\right)^{1/2}}{\Gamma(1/2)} \cdot y^{1/2-1} \cdot e^{-\frac{1}{2}y} \cdot \mathbb{1}_{\{y \geq 0\}}$$

   further solidifying our believes that $Y \sim \text{Gamma}(1/2,\ 1/2)$.

   (c) Suppose $T_i \overset{\text{i.i.d.}}{\sim} \chi^2$ for $i = 1, \cdots, n$. Identify the distribution of $W := \sum_{i=1}^n T_i$. **Hint:** One of the results from Discussion Worksheet 7 might be useful here. **By the Way:** The distribution of W is called the **χ^2 distribution with n degrees of freedom**, and is notated $W \sim \chi_n^2$.

   Solution: By Problem 1(a) on Worksheet 07, we know that the sum of independent Gamma random variables with the same rate parameter is also distributed as a Gamma random variable: i.e. if $X \sim \text{Gamma}(r, \lambda)$ and $Y \sim \text{Gamma}(s, \lambda)$ with $X \perp Y$ then $(X + Y) \sim \text{Gamma}(r + s,\ \lambda)$. In fact, this generalizes:

   $$X_i \overset{\text{i.i.d.}}{\sim} \text{Gamma}(r_i, \lambda) \implies \left(\sum_{i=1}^n X_i\right) \sim \text{Gamma}\left(\sum_{i=1}^n r_i,\ \lambda\right)$$

Therefore, since $T_i \overset{\text{i.i.d.}}{\sim} \text{Gamma}(1/2,\ 1/2)$ by part (b), we have

$$W := \sum_{i=1}^{n} T_i \sim \text{Gamma}\left(\sum_{i=1}^{n} \frac{1}{2},\ \frac{1}{2} \right) \sim \boxed{\text{Gamma}\left(\frac{n}{2},\ \frac{1}{2} \right)}$$

(d) Now, suppose $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$ for $i = 1, \cdots, n$. Define

$$S := \frac{1}{n} \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$$

Compute $\mathbb{E}[S]$ and $\text{Var}(S)$.

Solution: Recall that

$$\left(\frac{X_i - \mu}{\sigma} \right) \sim \mathcal{N}(0, 1)$$

[this is just our familiar Standardization result for the normal distribution]. Therefore,

$$\left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2$$

by the result we proved in the Transformations lecture. Therefore,

$$\left[\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \right] \sim \chi^n \sim \text{Gamma}\left(\frac{n}{2},\ \frac{1}{2} \right)$$

by part (c) above. Additionally, we have seen [on Problem 1(b) of Worksheet 5] that if $U \sim \text{Gamma}(r, \lambda)$ then $(cU) \sim \text{Gamma}(r,\ \lambda/c)$ for any positive constant c. Therefore,

$$S := \left(\frac{1}{n} \right) \cdot \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \text{Gamma}\left(\frac{n}{2},\ \frac{\left(\frac{1}{2}\right)}{\left(\frac{1}{n}\right)} \right) \sim \boxed{\text{Gamma}\left(\frac{n}{2},\ \frac{n}{2} \right)}$$

Since we know that $U \sim \text{Gamma}(r, \lambda) \implies \mathbb{E}[U] = r/\lambda$ and $\text{Var}(U) = r/(\lambda^2)$, we have

$$\mathbb{E}[S] = \frac{\left(\frac{n}{2}\right)}{\left(\frac{n}{2}\right)} = \boxed{1}; \quad \text{Var}(S) = \frac{\left(\frac{n}{2}\right)}{\left(\frac{n}{2}\right)^2} = \boxed{\frac{2}{n}}$$

2. Recall the notion of **runs**, discussed in lecture: a run, in the context of tossing a $p-$coin n times, refers to a string of consecutive heads or tails. For instance, in the outcome

$$H\ H\ H\ T\ T\ H\ T\ H\ T\ T$$

there are 6 runs:

$$H\ H\ H\ \mid T\ T \mid\ H \mid\ T \mid\ H \mid\ T\ T$$

Let X denote the number of runs in n tosses of a $p-$coin. We will also make the simplifying assumption that n is even.

(a) What is the state space of X?

Solution: As was mentioned in lecture, the smallest value X can take is 1. Analogously, the largest value X can take is n (alternating heads/tails), meaning

$$S_X = \{1, 2, \cdots, n\}$$

(b) Compute $\mathbb{P}(X = 1)$.

Solution: The event $\{X = 1\}$ corresponds to either all heads or all tails. Thus,

$$\mathbb{P}(X = 1) = p^n + q^n$$

(c) Compute $\mathbb{P}(X = n)$.

Solution: If the first toss landed heads, then the event $\{X = n\}$ corresponds to the sequence

$$HTHTHT \cdots HT$$

which has probability $p^{\frac{n}{2}} q^{\frac{n}{2}}$ of occurring. Similarly, if the first toss landed tails, then the event $\{X = n\}$ corresponds to the sequence

$$THTHTH \cdots TH$$

which also has probability $p^{\frac{n}{2}} q^{\frac{n}{2}}$. Therefore,

$$\mathbb{P}(X = n) = 2p^{\frac{n}{2}} q^{\frac{n}{2}}$$

It turns out that the PMF doesn't have a nice simple expression. Nonetheless, we can actually find a relatively simple expression for $\mathbb{P}(X = 2)$, which is what we will work toward in the next few parts.

(d) Suppose the first toss lands heads. Compute the probability of exactly 2 runs in this case.

Solution: Let's list out a few of the possible outcomes:

$$\begin{gathered} HH \cdots HHT \\ HH \cdots HTT \\ HH \cdots TTT \\ \vdots \\ HT \cdots TTT \end{gathered}$$

Adding up the probabilities of these outcomes yields

$$\begin{aligned} \text{answer} &= p^{n-1}q + p^{n-2}q^2 + \cdots + pq^{n-1} \\ &= \sum_{j=1}^{n-1} p^{n-j} q^j = p^n \sum_{j=1}^{n-1} \left(\frac{q}{p} \right)^j \\ &= p^n \times \frac{\left(\frac{q}{p} \right) - \left(\frac{q}{p} \right)^n}{1 - \left(\frac{q}{p} \right)} = \frac{qp^{n-1} - q^n}{1 - \left(\frac{q}{p} \right)} = \frac{p^n q - pq^n}{p - q} \end{aligned}$$

(e) Now, suppose the first toss lands tails. Compute the probability of exactly 2 runs in this case.

Solution: Let's list out a few of the possible outcomes:

$$\begin{gathered} TT \cdots TTH \\ TT \cdots THH \\ TT \cdots HHH \\ \vdots \\ TH \cdots HHH \end{gathered}$$

Adding up the probabilities of these outcomes yields

$$\begin{aligned} \text{answer} &= q^{n-1}p + q^{n-2}p^2 + \cdots + qp^{n-1} \\ &= \sum_{j=1}^{n-1} q^{n-j}p^j = q^n \sum_{j=1}^{n-1} \left(\frac{p}{q}\right)^j \\ &= q^n \times \frac{\left(\frac{p}{q}\right) - \left(\frac{p}{q}\right)^n}{1 - \left(\frac{p}{q}\right)} = \frac{pq^{n-1} - p^n}{1 - \left(\frac{p}{q}\right)} = \boxed{\frac{pq^n - p^n q}{q - p}} \end{aligned}$$

(f) Combine your two cases above to conclude

$$\mathbb{P}(X = 2) = 2 \cdot \frac{p^n q - pq^n}{p - q}$$

Solution: We need only to add our answers from parts (d) and (e) together:

$$\begin{aligned} \mathbb{P}(X = 2) &= \frac{p^n q - pq^n}{p - q} + \frac{pq^n - p^n q}{q - p} \\ &= \frac{p^n q - pq^n}{p - q} - \frac{pq^n - p^n q}{p - q} \\ &= \frac{p^n q - pq^n - pq^n + p^n q}{p - q} = \frac{2p^n q - 2pq^n}{p - q} = 2 \cdot \frac{p^n q - pq^n}{p - q} \end{aligned}$$

3. Given a collection of random variables $\{X_i\}_{i=1}^n$, we define the **sample mean** to be

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$$

For the purposes of this problem, we will assume that the X_i's are i.i.d. with common mean μ and common variance σ^2; i.e. $\mathbb{E}[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$ for all $i = 1, \cdots, n$, and the X_i's are all independent.

(a) Compute $\mathbb{E}[\bar{X}_n]$ as a function of μ and n.

Solution: By the Linearity of Expectation,

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n \mu = \frac{1}{\cancel{n}}(\cancel{n}\mu) = \boxed{\mu}$$

(b) Compute $\text{Var}(\bar{X}_n)$ as a function of σ^2 and n.

Solution: We have

$$\text{Var}(\bar{X}_n) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2}\sum_{i=1}^n \sigma = \frac{1}{n^2}\cdot(n\sigma^2) = \boxed{\frac{\sigma^2}{n}}$$

where we have utilized the independence of the X_i's to interchange the summation and the variance operator.

(c) If $X_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, what is the distribution of $\bar{X}_n$? Confirm that your answers to parts (a) and (b) are consistent with your answer to this part.

Solution: We know that the sum of i.i.d. normally distributed random variables is also normally distributed. Hence,

$$\boxed{\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)}$$

(d) If $X_i \overset{\text{i.i.d.}}{\sim} \text{Gamma}(r, \lambda)$, what is the distribution of $\bar{X}_n$? Confirm that your answers to parts (a) and (b) are consistent with your answer to this part.

Solution: We have

$$\left(\sum_{i=1}^n X_i\right) \sim \text{Gamma}(nr, \lambda)$$
$$\implies \frac{1}{n}\left(\sum_{i=1}^n X_i\right) \sim \boxed{\text{Gamma}\,(nr,\ n\lambda)}$$

To check consistency with parts (a) and (b),

$$\mathbb{E}[\bar{X}_n] = \frac{nr}{n\lambda} = \frac{r}{\lambda} = \mathbb{E}[X_i]\ \checkmark$$
$$\text{Var}(\bar{X}_n) = \frac{nr}{(n\lambda)^2} = \frac{r}{n\lambda^2} = \frac{\left(\frac{r}{\lambda^2}\right)}{n} = \frac{\text{Var}(X_i)}{n}\ \checkmark$$

4. The following parts are unrelated.

(a) Let X be a continuous random variable with MGF given by

$$M_X(t) = \begin{cases} (1-t)^{-3/2} & \text{if } t < 1 \\ \infty & \text{otherwise} \end{cases}$$

Find $\mathbb{E}[X]$ and $\text{Var}(X)$. **DO NOT** simply recognize the distribution of X; this part is designed to give you practice with differentiation!

Solution: We fix a $t < 1$ and then differentiate the MGF:

$$M_X'(t) = \frac{3}{2}(1-t)^{-5/2}$$

$$\implies \mathbb{E}[X] = M_X'(0) = \frac{3}{2}$$

$$M_X''(t) = \frac{3}{2} \cdot \frac{5}{2} \cdot (1-t)^{-7/2}$$

$$\implies \mathbb{E}[X^2] = M_X''(0) = \frac{3}{2} \cdot \frac{5}{2} = \frac{15}{4}$$

$$\implies \text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{15}{4} - \left(\frac{3}{2}\right)^2 = \frac{6}{4} = \frac{3}{2}$$

(b) Let X be a discrete random variable with MGF (Moment-Generating Function) given by

$$M_X(t) = (0.3 + 0.7e^{2t})^8$$

Find $p_X(k)$, the probability mass function (p.m.f.) of X.

Solution: First note: If $Y \sim \text{Bin}(n, p)$ then

$$M_Y(t) = (1 - p + pe^t)^n \quad \forall t \in \mathbb{R}$$

Therefore, if $Y \sim \text{Bin}(8, 0.7)$ we have

$$M_Y(t) = (0.3 + 0.7e^t)^8 \quad \forall t \in \mathbb{R}$$

Now, this is very close to the MGF of X, but not quite. Now, we also know that in general

$$M_{aY+b}(t) = e^{bt} M_X(at)$$

So, if $X = 2Y$ then

$$M_X(t) = M_{2Y}(t) = M_Y(2t) = (0.3 + 0.7e^{2t})^8 \quad \forall t \in \mathbb{R}$$

which *is* the MGF of X! Therefore, we take $Y \sim \text{Bin}(8, 0.7)$ so that $X = 2Y$ and

$$p_X(k) = \mathbb{P}(X = k) = \mathbb{P}(2Y = k) = \mathbb{P}(Y = k/2) = \binom{8}{k/2} (0.7)^{k/2} (0.3)^{8-k/2} \cdot \mathbb{1}_{\{k \in \{0,2,4,6,\cdots,16\}\}}$$