# STA261 (Summer 2022) - Assignment 0

These problems are meant to refresh your STA257 skills. They are *not* to be handed in.

---

1. (a) Let $X \sim \mathcal{N}\left(0, \sigma^2\right)$. Show that $\mathbb{E}\left[X^{2k+1}\right] = 0$ for any $k \in \mathbb{N}$.

   (b) Go a bit further and show that this is true for *any* continuous distribution which is symmetric about zero (i.e., its pdf satisfies $f_X(x) = f_X(-x)$ for any $x \in \mathbb{R}$), provided all of its moments are finite of course. In other words, if a distribution is symmetric about zero, then *all* of its odd moments must vanish.

2. Let $X \sim \text{Poisson}\,(\lambda)$ and let $h : \mathbb{R} \to \mathbb{R}$ be any function such that $\mathbb{E}\left[h(X)\right]$ is finite. Prove that $\mathbb{E}\left[\lambda \cdot h(X)\right] = \mathbb{E}\left[X \cdot h(X-1)\right]$.

3. Let $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ and let $g : \mathbb{R} \to \mathbb{R}$ be any differentiable function that's nice enough to satisfy $\lim_{x\to\infty} g(x) \cdot e^{-x^2} = 0$. Prove that $\mathbb{E}\left[g(X) \cdot (X - \mu)\right] = \sigma^2 \cdot \mathbb{E}\left[g'(X)\right]$.

4. For any set of univariate random variables $X_1, X_2, \ldots, X_n$, the **order statistics** are the $X_i$'s placed in ascending order, which are notated as $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$. Thus the **sample minimum** $X_{(1)} = \min\{X_1, \ldots, X_n\}$ and the **sample maximum** $X_{(n)} = \max\{X_1, \ldots, X_n\}$.

   In STA257, you may have learned that if $X_1, X_2, \ldots, X_n$ are an independent sample from a continuous distribution with pdf $f_X$ and cdf $F_X$, then $f_{X_{(1)}}(x) = nf_X(x)(1 - F_X(x))^{n-1}$ and $f_{X_{(n)}}(x) = nf_X(x)F_X(x)^{n-1}$. Let's generalize those formulas by finding the pdf of $X_{(j)}$, for any $1 \le j \le n$.

   (a) Let $h > 0$ be nice and small. Explain why

   $$\begin{aligned}&\mathbb{P}\left(X_{(j)} \in [x, x+h]\right) \\ &= \mathbb{P}\left(\text{One of the } X_i\text{'s is in } [x, x+h] \text{and exactly } j-1 \text{ of the others are } < x\right).\end{aligned}$$

   (b) Show that the probability on the right is equal to

   $$n \cdot \mathbb{P}\left(X_1 \in [x, x+h]\right) \cdot \mathbb{P}\left(\text{exactly } j-1 \text{ of the others are } < x\right).$$

   (c) Think Binomially and show that

   $$\mathbb{P}\left(\text{exactly } j-1 \text{ of the others are } < x\right) = \binom{n-1}{j-1} F_X(x)^{j-1}(1 - F_X(x))^{n-j}.$$

   (d) Put the pieces together, divide both sides by $h$, and take the limit as $h \to 0$ to get

   $$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x)F_X(x)^{j-1}(1 - F_X(x))^{n-j}.$$

5. Let $X_1, X_2, \ldots, X_n$ be independent $\text{Unif}\,(0,1)$ random variables. Show $X_{(j)} \sim \text{Beta}\,(j, n-j+1)$, and use that fact to find $\mathbb{E}\left[X_{(j)}\right]$ and $\text{Var}\left(X_{(j)}\right)$.

6. Let $U_1, U_2, \ldots$ be independent $\text{Unif}\,(0,1)$ random variables. Let $X$ be a random variable with distribution

   $$\mathbb{P}\left(X = x\right) = \frac{c}{x!}, \quad x = 1, 2, 3, \ldots,$$

   for some $c \in \mathbb{R}$. Find the value of $c$, and then find the distribution of $Z = \min\{U_1, U_2, \ldots, U_X\}$. That's the minimum of a random number of $U_i$'s, so you'll have to do some kind of conditioning.

7. Suppose you repeatedly draw independent standard uniform random variables and add them together. What's the expected number of draws you need for the sum to exceed 1? Let's answer that.

   (a) Let $U_1, U_2, \dots, U_n$ be independent $\text{Unif}(0,1)$ random variables, and let $S_n = \sum_{i=1}^n U_i$. Prove that $\mathbb{P}(S_k \leq t) = t^k/k!$ for $t \in (0,1)$ using mathematical induction.[1]

   (b) Let $N = \min\{k : S_k > 1\}$. Argue that $\mathbb{P}(N = n) = \mathbb{P}(S_{n-1} < 1) - \mathbb{P}(S_n < 1)$.

   (c) Use that to evaluate $\mathbb{E}[N]$. Think about where your summation starts!

8. Find $\mathbb{P}\left(X > \sqrt{Y}\right)$, if $X$ and $Y$ are jointly distributed according to

$$f_{X,Y}(x,y) = x + y, \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1.$$

9. Let $B$ and $C$ be independent $\text{Unif}(0,1)$ random variables. Find the probability that the random quadratic $x^2 + Bx + C$ has a real root.

10. Let $Y$ be a random variable whose first two moments exist. Hypothesize which $x \in \mathbb{R}$ minimizes $\mathbb{E}\left[(Y-x)^2\right]$, and then prove it.

11. Let $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\rho)$ be independent. Find the conditional distribution of $X \mid (X + Y = n)$.

12. Let $X \sim \text{Gamma}(\lambda, 1)$ and $Y \sim \text{Gamma}(\rho, 1)$ be independent. Name the distributions of $G = X + Y$ and $B = X/(X+Y)$, and show they're independent. Don't try to start by finding the marginals – instead, go straight for the joint distribution of $(G, B)$ and see what pops out.

13. Let $X$ and $Y$ be independent $\mathcal{N}(0,1)$ random variables.

    (a) Let $R = \sqrt{X^2 + Y^2}$ and $\Theta = \arctan\left(\frac{Y}{X}\right)$. Name the distributions of $R^2$ and $\Theta$, and show they're independent. If your trig is rusty, remember that $\tan(x) = \sin(x)/\cos(x)$ and $\sin^2(x) + \cos^2(x) = 1$.

    (b) Use your work to show that if $U_1$ and $U_2$ are independent $\text{Unif}(0,1)$ random variables, then $X \overset{d}{=} \sqrt{-2\log(U_1)}\cos(2\pi U_2)$ and $Y \overset{d}{=} \sqrt{-2\log(U_1)}\sin(2\pi U_2)$.

    (c) If I give you only a pocket calculator and two independent draws from the $\text{Unif}(0,1)$ distribution, explain how you can give me back independent draws from the $\mathcal{N}\left(\mu_1, \sigma_1^2\right)$ distribution and the $\mathcal{N}\left(\mu_2, \sigma_2^2\right)$ distribution.

14. Let $X_1, X_2$ and $X_3$ be uncorrelated random variables, all with expectation $\mu$ and variance $\sigma^2$. Find expressions for $\text{Cov}(X_1 + X_2, X_2 + X_3)$ and $\text{Cov}(X_1 + X_2, X_1 - X_2)$ in terms of $\mu$ and $\sigma^2$.

15. Let $X_1, X_2, \dots, X_n$ be independent random variables with $\mathbb{E}[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. Define the **sample mean** $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and the **sample variance** $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Prove that $\mathbb{E}\left[\bar{X}_n\right] = \mu$ and $\mathbb{E}\left[S_n^2\right] = \sigma^2$. You can make life easier by writing $X_i - \bar{X} = (X_i - \mu) - (\bar{X} - \mu)$.

16. In the same setting as above, show that the sample variance satisfies

$$S_n^2 = \frac{1}{2n(n-1)} \sum_{i,j=1}^n (X_i - X_j)^2.$$

    Interpret this result.

[1] If you don't know what this is, just follow these steps: first prove the result holds when $k = 1$. Then *assume* the result holds for any $k \in \mathbb{N}$, and show that this implies the result must also hold for $k + 1$. The *principle of mathematical induction* says that if you've done that, then you've proven the result holds for all $k \in \mathbb{N}$.

17. Let $A$ be an $n \times n$ matrix whose entries are independent $\mathcal{N}(0,1)$ random variables. Let $B = \frac{1}{2}(A + A^T)$, which you might notice is symmetric. What is the joint pdf of the $n(n+1)/2$ entries in the upper triangle of $B$? This has matrices in it, but it doesn't need any linear algebra; if you remember what the transpose of a matrix is, you can do this! If you're looking for a name for your pdf, you can call it $f_{B_{11},B_{12},\ldots,B_{nn}}(b_{11}, b_{12}, \ldots, b_{nn})$.

18. What's the probability that an unbiased coin lands on heads 500 times in 1000 flips, rounded to five decimal places? You know that the exact answer is $\binom{1000}{500}0.5^{1000}$, but good luck trying to evaluate that on a calculator – you'll either end up with numerical underflow or overflow. You might think to calculate the log of that and then exponentiate it after – that will definitely help with the $0.5^{1000}$ part, but you'll still have to deal with $\log(1000!) - 2\log(500!)$, and you just can't evaluate either of those factorials directly. You may have heard of *Stirling's formula*, which gives an approximation of the factorial function. With a bit of hand-waving, we'll derive a version of it here.

    (a) Let $X_1, X_2, \ldots, X_n$ be independent $\text{Exp}(1)$ random variables. Using mgfs (or anything else), show that $\sum_{i=1}^n X_i \sim \text{Gamma}(n, 1)$.

    (b) Fix $x \in \mathbb{R}$. Explain why we can write

    $$\frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}\left(\frac{\bar{X}_n - 1}{1/\sqrt{n}} \leq x\right) \approx \phi(x)$$

    when $n$ is large, where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is the standard Normal pdf.

    (c) Carry out the differentiation on the left-hand side, via a $u$-substitution and the FTC.

    (d) Take $x \to 0$ on both sides and rearrange a bit to get

    $$n! \approx \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n},$$

    which is Stirling's formula.

    (e) Approximate (to five decimal places) the probability that an unbiased coin lands on heads 500 times in 1000 flips. I get 0.02523...

19. Fix some $n \in \mathbb{N}$ with $n > 1$. Prove that if I give you some fixed $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, you can give me $x_1, x_2, \ldots, x_n \in \mathbb{R}$ such that

    $$\bar{x} := \frac{1}{n}\sum_{i=1}^n x_i = \mu$$

    and

    $$\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2 = \sigma^2.$$

    What — if any — are some statistical implications of this?

    Hint: start with $n = 2$, and let you'll get an explicit form for $x_1$ and $x_2$. Use those to take a guess at the case for general $2n$, and prove that it gives you what you want. For odd $n$, add an appropriate $x_{2n+1}$ to the $2n$ case.

20. In STA257, you learned *Chebyshev's inequality*, a corollary of which says that if $\mathbb{E}[X] = \mu$ and $\text{Var}(X) = \sigma^2$, then $\mathbb{P}(|X - \mu| > \lambda) \leq \sigma^2/\lambda^2$ for any $\lambda > 0$. This is the most basic example of a *concentration inequality*, so named because it essentially says that random variables with finite moments tend to "concentrate" around their means — in this case, the probability that $X$ is at distance at least $x$ away from $\mu$ decays like $1/x^2$. It turns out that Chebyshev's inequality is rather weak, and for sums of nice independent random variables, we can obtain much stronger concentration.

(a) First show that Chebyshev's inequality is tight (i.e., equality holds for some random variable $X$ and some $\lambda > 0$). The easiest example is discrete — try and construct $X$ so that it gives you what you need.

(b) Let $X_i \sim \text{Bernoulli}(p_i)$ be independent for $i = 1, \ldots, n$. Let $X = \sum_{i=1}^n X_i$ and $\mu = \sum_{i=1}^n p_i$.

i. Let $M_X(t)$ be the mgf of $X$. Use the fact that $1+x \leq e^x$ to show that $M_X(t) \leq e^{(e^t-1)\mu}$.

ii. Use Markov's inequality and the inequality above to show that for any $\delta > 0$ and any $t \in \mathbb{R}$,

$$\mathbb{P}(X > (1+\delta)\mu) \leq \left( \frac{e^{(e^t-1)}}{e^{(1+\delta)t}} \right)^{\mu}.$$

iii. Minimize the right-hand side in $t$ to show that

$$\mathbb{P}(X > (1+\delta)\mu) \leq \left( e^{\delta-(1+\delta)\log(1+\delta)} \right)^{\mu}.$$

iv. Using a Taylor expansion of $\log(1+\delta)$, show that $-(1+\delta)\log(1+\delta) \leq -\delta^2/3 - \delta$ for $\delta < 1$ and conclude that

$$\mathbb{P}(X > (1+\delta)\mu) \leq e^{-\delta^2\mu/3}.$$

How does this compare to the kind of bound you'd get with Chebyshev?