STAT430 Exam 1.

Name: KEY

____

Time Limit (strict): 50 Minutes

Closed book, closed notes, no calculator. No electronic devices of any kind. Silence phones and put them away.

##Keep your eyes on your own paper and show your work.##

Honor Code: I will not give, receive, or use any unauthorized assistance on this exam.

Your signature:

____

1. Suppose $Z_1, Z_2, \ldots, Z_n$ are independent and identically distributed (iid) $N(0,1)$. Let $\bar{Z} = (1/n)\sum_{i=1}^n Z_i$ denote the usual sample mean, and let $S^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2/(n-1)$ denote the usual sample variance. Let t_ν denote the t distribution with ν degrees of freedom (df), χ^2_ν denote the chi-square distribution with ν df, and $\mathcal{F}_{\nu_1,\nu_2}$ denote the F distribution with ν_1 numerator df and ν_2 denominator df.

- (a). Circle **TRUE** or **FALSE**:

$$\frac{Z_1}{\sqrt{\sum_{i=1}^n Z_i^2/n}} \sim t_n.$$

FALSE. Pay attention to whether the numerator and denominator are independent. In this case they are not. For this to be distributed as t, we need a standard normal random variable in the numerator and the square root of an independent χ^2 random variable divided by its degrees of freedom in the denominator.

- (b). Circle **TRUE** or **FALSE**:

$$\frac{(Z_1^2 + Z_2^2)/2}{\sum_{i=3}^n Z_i^2/(n-2)} \sim \mathcal{F}_{2,n-2}.$$

TRUE. The numerator is definitely independent of the denominator, because they involve none of the same random variables. Since we have a χ^2_2 random variable divided by its degrees of freedom in the numerator, and an independent χ^2_{n-2} random variable divided by its degrees of freedom in the denominator, this whole thing is distributed as and $\mathcal{F}_{2,n-2}$

- (c). Circle **TRUE** or **FALSE**:

$$\sum_{i=1}^n (Z_i - \bar{Z})^2 \sim \chi^2_{n-1}.$$

TRUE. (I got this wrong originally). Notice that this is equal to $\frac{(n-1)}{\sigma^2}S^2$, where $\sigma^2 = 1$. This is distributed χ^2_{n-1}. This is something that you should remember.

- (d). Circle **TRUE** or **FALSE**:

$$\frac{(Z_4^2 + Z_5^2)/(2-1)}{(Z_1^2 + Z_2^2 + Z_3^2)/(3-1)} \sim \mathcal{F}_{1,2}.$$

FALSE. The sums of standard normal random variables are respectively χ^2_2 and χ^2_3 in the numerator and denominator. And they are independent, but dividing by random degrees of freedom doesn't produce an $\mathcal{F}$ distribution with those degrees of freedom. At least I don't think it does. As far as I'm aware, this doesn't correspond to a random variable with a distribution we should know.

- (e). Give a complete specification of the distribution of $\bar{Z}$. Justify your result.

$\bar{Z} \sim N(0, 1/n)$

$$E(\bar{Z}) = E\left(\frac{1}{n}\sum_{i=1}^n Z_i\right) = \frac{1}{n}\sum_{i=1}^n E(Z_i) = \frac{1}{n}nE(Z_1) = 0,$$

and

$$V(\bar{Z}) = V\left(\frac{1}{n}\sum_{i=1}^n Z_i\right) = \frac{1}{n^2}\sum_{i=1}^n V(Z_i) = \frac{1}{n^2}nV(Z_1) = \frac{1}{n}.$$

- (f). Give a complete specification of the distribution of $n\bar{Z}^2/S^2$. Justify your result.

$n\bar{Z}^2/S^2 \sim F_{1,n-1}$ From the above question, we should know that $\bar{Z} \sim N(0, 1/n)$, this means that $\frac{\bar{Z}}{1/\sqrt{n}} = \sqrt{n}\bar{Z} \sim N(0,1)$, so $n\bar{Z}^2 \sim \chi^2_1$. We should also know that $(n-1)S^2 \sim \chi^2_{n-1}$. So if we multiply and divide the denominator by $n-1$, we get

$$\frac{(n\bar{Z}^2)/1}{((n-1)S^2)/(n-1)}.$$

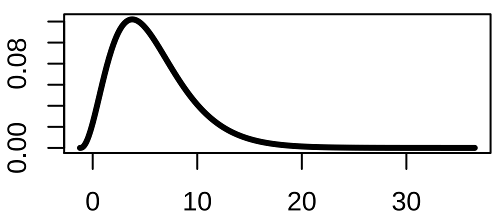
$\bar{Z}$ and S^2 are independent, so we have a χ^2_1 random variable divided by its degrees of freedom in the numerator, and an independent χ^2_{n-1} random variable divided by its degrees of freedom in the denominator. This corresponds to a random variable with a $\mathcal{F}_{1,n-1}$ distribution.

- (g). **Circle one.** The person who originally derived the t_ν distribution worked for a: (A) pharmaceutical company; (B) hospital; (C) brewery; or (D) university.

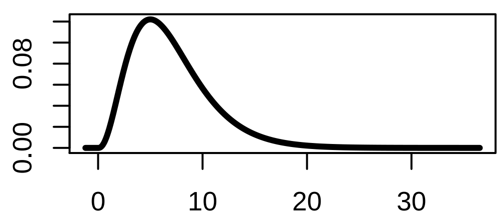
Brewery (Guiness, Specifically)

---

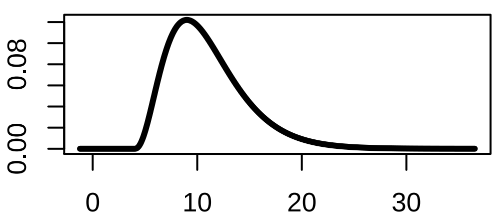
2. Which of the following curves is the probability density function of a χ^2_7, a chi-square random variable with 7 degrees of freedom? **Circle the correct Option.**



Option A



Option B



Option C

A random variable $\sim \chi^2_7$ distribution has an expected value equal to its degrees of freedom (7). Since the distribution is also right-skewed, we know that the expected value will be greater than the median value. Also, a χ^2 random variable's support is for $[0, \infty)$. Notice that option A has a tiny bit of non-zero density for values less than zero. I would go with option B.

3. Suppose $Y \sim \text{Binomial}(100, 0.5)$ and we want to approximate the following probability,

$$P\,(50 < Y \leq 100)\,,$$

using the central limit theorem. Choose the best answer, where $Z \sim N(0,1)$.

$$\begin{array}{ll}
\text{(A)} & P\left(\dfrac{50-50}{0.5} \leq Z \leq \dfrac{100-50}{0.5}\right) \\
\text{(B)} & P\left(\dfrac{50-50}{0.25} \leq Z \leq \dfrac{100-50}{0.25}\right) \\
\text{(C)} & P\left(\dfrac{50-50}{25} \leq Z \leq \dfrac{100-50}{25}\right) \\
\text{(D)} & P\left(\dfrac{50-50}{5} \leq Z \leq \dfrac{100-50}{5}\right)
\end{array}$$

By the CLT, we know that

$$\frac{Y - n\mu}{\sqrt{n}\sigma} \overset{\text{approx}}{\sim} N(0,1)$$

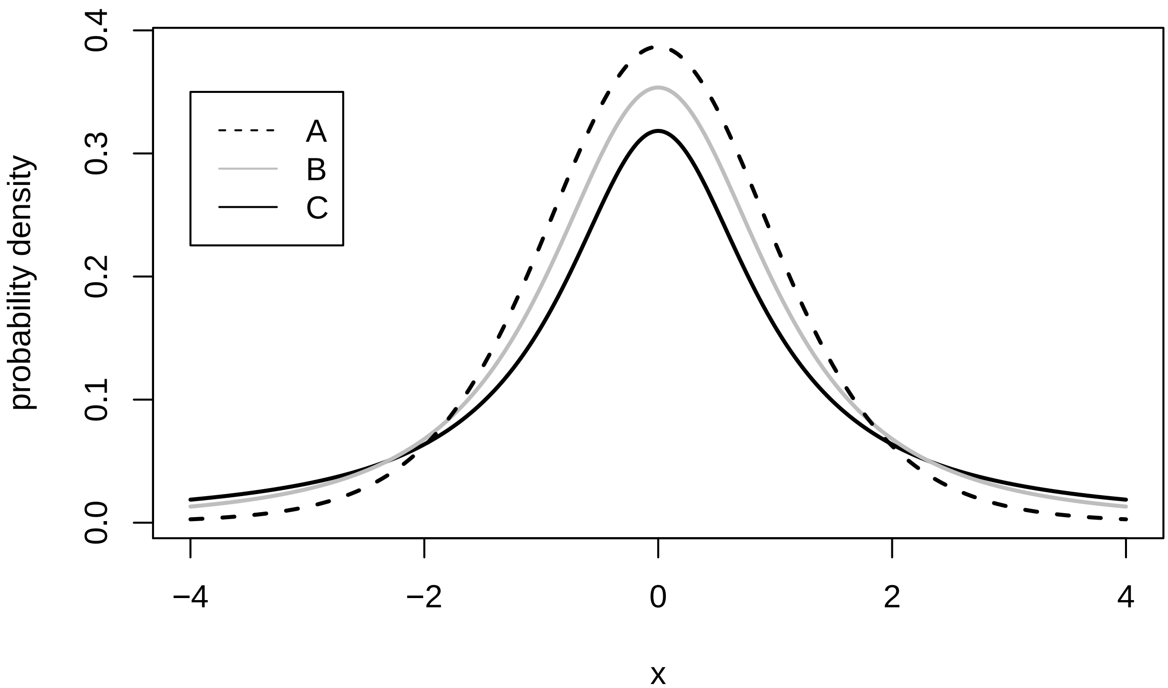
So we can use this to estimate the probability with

$$P\left(\frac{50-50}{5} \leq Z \leq \frac{100-50}{5}\right)$$

Where $Z \sim N(0,1)$.

---

4. The three probability density functions (pdf's) in the following graph are for three different t distributions. Among the three, which one (A, B or C) corresponds to the t distribution with smallest degrees of freeedom? Explain briefly.



C - honestly, I looked at the Wikipedia page for the t distribution. The bigger the degrees of freedom are, the higher the peak is. This makes sense, because if we're thinking about the construction of a t random variable using iid standard normal random variables, the bigger our sample size is, the more concentrated things will be around the mean.

---

5. If X is a random variable with $E(X) = \mu_X$ and $V(X) = \sigma_X^2$, Y is a random variable with $E(Y) = \mu_Y$ and $V(Y) = \sigma_Y^2$, and X and Y are independent, find $V(XY)$.

If X and Y are independent, then $E(f(X)g(Y)) = E(f(X))E(g(Y))$.

$$\begin{aligned}
V(XY) &= E((XY)^2) - (E(XY))^2 \\
&= E(X^2)E(Y^2) - (E(X))^2(E(Y))^2 \\
&= (V(X) + (E(X))^2)(V(Y) + (E(Y))^2) - (E(X))^2(E(Y))^2 \\
&= (\sigma_X^2 + \mu_X^2)(\sigma_Y^2 + \mu_Y^2) - \mu_X^2\mu_Y^2
\end{aligned}$$

---

6. Suppose that $Y_1, \ldots, Y_n$ are iid according to the following pdf:

$$f(y) = \begin{cases} 1/4, & \theta - 1 \le y \le \theta + 3, \\ 0, & \text{otherwise}, \end{cases}$$

where θ is an unknown parameter. Consider $\hat{\theta}_1 = \bar{Y}$ as an estimator of the target parameter, θ.

- 6(a). Compute $B(\hat{\theta}_1)$, the bias of $\hat{\theta}_1$ as an estimator of θ.

$B(\hat{\theta}_1) = E(\hat{\theta}_1) - \theta$.

$$\begin{aligned} B(\hat{\theta}_1) &= E(\bar{Y}) - \theta = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) - \theta \\ &= E(Y_1) - \theta = \int_{\theta-1}^{\theta+3} \frac{y}{4} dy - \theta \\ &= \frac{y^2}{8}\Big|_{\theta-1}^{\theta+3} - \theta = \frac{(\theta+3)^2 - (\theta-1)^2}{8} - \theta \\ &= \frac{\theta^2 + 6\theta + 9 - \theta^2 + 2\theta - 1}{8} - \theta = \frac{8\theta + 8}{8} - \theta \\ &= \theta + 1 - \theta = 1 \end{aligned}$$

- 6(b). Show how to use $\hat{\theta}_1$ to construct a new, unbiased estimator $\hat{\theta}_2$ of θ.

Since $E(\hat{\theta}_1) = \theta + 1$, we can construct an unbiased estimator for θ by taking $\hat{\theta}_2 = \hat{\theta}_1 - 1$. Then

$$E(\hat{\theta}_2) = E(\hat{\theta}_1 - 1) = E(\hat{\theta}_1) - 1 = \theta + 1 - 1 = \theta$$

7(a). If $X_1 \sim \text{Poisson}(\lambda_1)$ independent of $X_2 \sim \text{Poisson}(\lambda_2)$, then $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$. The exact distribution of $W = \sum_{i=1}^{n} Y_i$, where $Y_1, \ldots, Y_n$ are iid Poisson(1), is therefore Poisson(n). Suppose $n = 100$. Use the Central Limit Theorem (explaining why it is applicable) to approximate $P\left(W \leq 120\right)$.

$$\frac{W - n\lambda}{\sqrt{n\lambda}} \overset{\text{approx}}{\sim} N(0,1).$$

So we can approximate the true value of $P(W \leq 120)$ using a standard normal distribution and finding

$$P\left(W \leq 120\right) \approx \Phi(120),$$

Where

$$\Phi(120) = \int_{-\infty}^{120} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}x^2} dx.$$

7(b). The `R` code below computes various probabilities. Choose the exact value of $P\left(W \leq 120\right)$, and compare to your answer in 7(a).

```
dpois((120 - 100) / sqrt(100), lambda = 1)
```

```
## [1] 0.1839397
```

```
1 - ppois(120, lambda = 100)
```

```
## [1] 0.02266933
```

```
dpois(120, lambda = 100)
```

```
## [1] 0.005561065
```

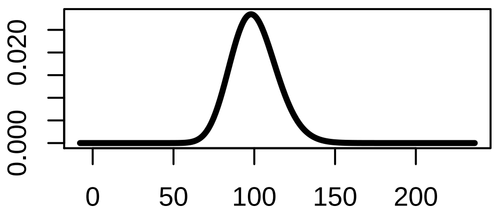
```
sum(dpois(0:120, lambda = 100))
```

```
## [1] 0.9773307
```

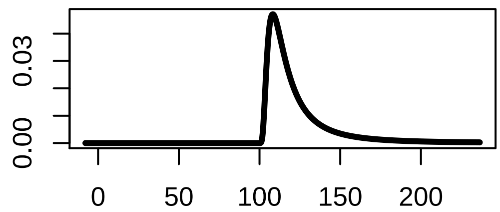
The last one is correct. `dpois()` does not give the probability of being less than given quantile, it returns the value of the PMF (for discrete, this is the probability that the random variable = the given quantile), however, if we sum up all the values of the PMF for values from 0 to 120, we will get the cumulative probability that the random variable will be less than or equal to 120.

---

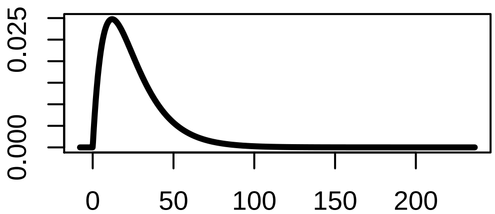
8. Which of the following curves is the probability density function of a χ^2_{100}, a chi-square random variable with 100 degrees of freedom? **Circle the correct Option.**



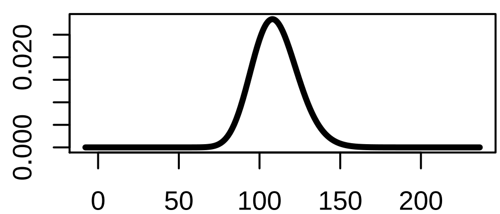
Option A



Option B



Option C



Option D

A χ^2_{100} random variable is the sum of 100 χ^2_1 iid random variables. Sound familiar? It should. The CLT will kick in and the distribution will start looking normal. The average is 100, (the degrees of freedom), and a χ^2 random variable still only has support on the non-negative part of the real line. Even though it should look pretty normal, it's still a skewed distribution and the median will be slightly lower than the mean. Option A is the right one I think.

---

9. Suppose $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed (iid) $N(\mu, \sigma^2)$ and let S^2 denote the usual sample variance. In clqww, we showed that $V(S^2) = 2\sigma^4/(n-1)$. What is $E\left(S^4\right) = E\left[(S^2)^2\right]$? What is the limit of $E(S^4)$ as $n \to \infty$?

$$
\begin{aligned}
E(S^4) &= E((S^2)^2) \\
&= V(S^2) + (E(S^2))^2 \\
&= \frac{2\sigma^4}{n-1} + \sigma^4
\end{aligned}
$$

$$
\lim_{n\to\infty}\left(\frac{2\sigma^4}{n-1} + \sigma^4\right) = 0 + \sigma^4 = \sigma^4
$$

$$P(Z \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\, dw$$

$$\Phi(-z) = 1 - \Phi(z)$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990