Math 10B with Professor Stankova
Worksheet, Discussion #29; Thursday, 5/9/2019
GSI name: Roy Zhao

1 Continuous Random Variables

1. **TRUE** False If the mean doesn't exist, then the standard deviation doesn't exist.

> **Solution:** The formula for the standard deviation requires the mean, so if the mean doesn't exist, then we can't talk about the standard deviation.

2. True **FALSE** If the mean exists, then the standard deviation exists.

> **Solution:** It is possible for the mean to exist but the standard deviation not to exist. For example, the distribution $\frac{1}{x^3}$ on $x \geq 1$ has the mean existing but the standard deviation not.

3. True **FALSE** CDFs are always continuous but PDFs don't have to be.

> **Solution:** Both can be non-continuous.

4. **TRUE** False If we know the CDF F, we find the probability $P(2 \leq X \leq 5)$ without calculating the PDF f.

> **Solution:** We have $P(2 \leq X \leq 5) = F(5) - F(2)$.

5. **TRUE** False The variance of a symmetric random variable centered at 0 is $\int_{-\infty}^{\infty} x^2 f(x)dx$.

> **Solution:** If it is symmetric and centered at 0, then $E[X] = 0$.

6. **TRUE** False If a is the median of a continuous random variable X with PDF f, then $\int_a^\infty f(x)dx = \frac{1}{2}$.

7. Let X be a random variable on a probability space Ω with a probability function P and let f be the PDF for X, F the CDF. Draw a picture of how all these variables interact and explain any special arrows that you have in your diagram.

Solution: P is a dashed line from Ω to $[0,1]$ because it takes in subsets of Ω. X is a solid arrow from Ω to $\mathbb{R}$ because it is a function that takes outcomes to a value. f is a solid arrow from $\mathbb{R}$ to $[0,\infty)$. F is a solid line from $\mathbb{R}$ to $[0,1]$ and integration is a dashed line from $\mathbb{R}$ to $[0,1]$.

8. Let $g(x) = \begin{cases} x^2 & -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$. Find c such that $f(x) = cg(x)$ is a PDF. Graph f and the CDF F. Find the mean and median of $f(x)$.

Solution: First we calculate

$$\int_{-\infty}^{\infty} g(x)dx = \int_{-1}^{1} x^2 dx = \frac{2}{3}.$$

Therefore, we must have that $\frac{2}{3}c = 1$ or $c = \frac{3}{2}$. The CDF is

$$F(x) = \begin{cases} 0 & x \le -1 \\ \int_{-1}^{x} 3/2t^2 dt & -1 \le x \le 1 \\ \int_{-1}^{1} 3/2t^2 dt & 1 \le x \end{cases} = \begin{cases} 0 & x \le -1 \\ \frac{x^3+1}{2} & -1 \le x \le 1 \\ 1 & 1 \le x \end{cases}.$$

The median is when $F(x) = 1/2$ or when $x^3 + 1 = 1$ which is $x = 0$. The mean is

$$\mu = \int_{-\infty}^{\infty} xf(x)dx = \int_{-1}^{1} 3/2x^3 dx = 0,$$

since $3/2x^3$ is an odd function. So again, the median and mean align.

9. Let $F(x) = \frac{x-1}{x+1}$ for $x \ge 1$ and 0 for $x \le 1$. Show that F is a CDF. Find the PDF associated with it and the probability that we choose a number between 1 and 2.

Solution: This is a CDF because it is continuous since $F(1) = 0$ and $\lim_{x\to\infty} F(x) = 1$ and F is non-decreasing. The PDF is

$$f(x) = \frac{d}{dx}F(x) = \frac{2}{(x+1)^2}$$

for $x \geq 1$ and 0 for $x \leq 1$. The probability that we choose a number between 1 and 2 is
$$\int_1^2 f(x)dx = F(2) - F(1) = \frac{1}{3}.$$

10. Let $f(x)$ be $-2x$ from $-1 \leq x \leq 0$ and 0 everywhere else. Find the standard deviation of this distribution.

Solution: First we find the mean as
$$\int_{-\infty}^{\infty} xf(x)dx = \int_{-1}^{0} -2x^2dx = -2/3x^3|_{-1}^{0} = \frac{-2}{3}.$$
Then, to find the variance, we take
$$\sigma^2 = \int_{-\infty}^{\infty}(x-(-2/3))^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 = \int_{-1}^{0} -2x^3dx - (-2/3)^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}.$$
Thus, $\sigma = \sqrt{1/18} = \sqrt{2}/6$.

11. Let $f(x) = e \cdot e^x$ for $x \leq -1$ and 0 otherwise. Find the standard deviation of this distribution.

Solution: First we need to find the mean of this distribution. The mean is
$$\int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{-1} x(e \cdot e^x)dx + \int_{-1}^{\infty} 0dx = e\int_{-\infty}^{-1} xe^x dx$$
$$= e(xe^x - e^x|_{-\infty}^{-1}) = e[(-e^{-1} - e^{-1}) - 0] = -2.$$
To find the standard deviation, we first find the variance and then take the square root. There are two ways to do this, the latter is a bit easier
$$\sigma^2 = \int_{-\infty}^{\infty}(x-(-2))^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 = \int_{-\infty}^{-1} x^2(e \cdot e^x)dx - 4$$
$$= e(x^2e^x - 2xe^x + 2e^x|_{-\infty}^{-1}) - 4 = e(e^{-1} + 2e^{-1} + 2e^{-1}) - 4 = 5 - 4 = 1.$$
So the standard deviation is $\sigma = 1$.

12. Prove all of the formulas for mean and variance for each of the random variables below:

Distribution	PDF	$E(X)$	Variance
Uniform	$f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2
Exponential	$f(x) = ce^{-cx}$	$\frac{1}{c}$	$\frac{1}{c^2}$
Laplacian	$f(x) = \frac{1}{2}e^{-\lvert x\rvert}$	0	2

2 CLT

13. True **FALSE** A smaller 95% confidence interval means that we are less sure about what the mean μ could be.

Solution: It means that we are more sure because we are 95% sure it is in a smaller range.

14. True **FALSE** The normalized random variable $Z = \frac{\bar{X}-\bar{\mu}}{\sigma/\sqrt{n}}$ is standard normally distributed.

Solution: It is approximately standard normally distributed.

15. Suppose that the height of women is distributed with an average height of 63 inches and a standard deviation of 10 inches. Taking a sample of 100 women, what is the probability that the average of the heights of these 100 women is between 62 and 64 inches?

Solution: The average height of 100 women will be approximately normally distributed with average 63 and standard deviation $10/\sqrt{100} = 1$. Therefore, $P(62 \leq \bar{X} \leq 64) = P(62 \leq \bar{X} \leq 63) + P(63 \leq \bar{X} \leq 64) = z(1) + z(1) = 2z(1)$.

16. Suppose the weight of newborns is distributed with an average weight of 8 ounces and a standard deviation of 1 ounce. Today, there were 25 babies born at the Berkeley hospital. What is the probability that the average weight of these newborns is less than 7.5 ounces?

Solution: The average weight of these babies will be approximately normally distributed with mean 8 and standard deviation $1/\sqrt{25} = 0.2$. The probability is $P(\bar{X} \leq 7.5) = 0.5 - P(7.5 \leq \bar{X} \leq 8) = 0.5 - z(2.5)$.

17. I have a (possibly biased) coin and flip it 100 times and get heads 90 times. What is the 95% confidence interval for p, the probability of flipping a heads?

> **Solution:** The estimator for $\hat{p} = \frac{90}{100}$ and $\hat{\sigma} = \sqrt{\hat{p}(1-\hat{p})} = \frac{3}{10}$. Then the 95% confidence interval is $(\hat{p} - 2\hat{\sigma}/\sqrt{n}, \hat{p} + 2\hat{\sigma}/\sqrt{n}) = (0.9 - 2(0.3)/10, 0.9 + 2(0.3)/10) = (0.9 - 0.06, 0.9 + 0.06) = (0.84, 0.96)$.

18. Assume the standard deviation of student heights is 5 inches. How large of a sample do you need to be 95% confident that the sample mean is within 1 inch of the population mean?

> **Solution:** We want to be between 1 of the population mean and hence our interval radius $2\sigma/\sqrt{n}$ must be equal to 1. Therefore, we have that $2\sigma/\sqrt{n} = 1$ and $\sigma = 5$ so $\sqrt{n} = 10$ and $n = 100$.

19. In a class of 25 students, the time that students spent on the midterm was 40 minutes with a standard deviation of 5 minutes. What is the 95% confidence interval for the average time taken on the midterm?

> **Solution:** We have that $\mu = 40, n = 25, \sigma = 5$. Plugging it into our formula gives $(40 - 2 \cdot 5/\sqrt{25}, 40 + 2 \cdot 5/\sqrt{25}) = (38, 42)$.

20. I have a loaded die and I think that it is more likely to be a 1 than normal. Suppose I roll it 100 times and get 1 25 times. What is the 95% confidence interval for p, the probability of getting a 1?

> **Solution:** The estimator for p is $\frac{25/100}{=}\frac{1}{4}$. Given this, the estimator for the standard deviation is $\hat{\sigma} = \sqrt{\hat{p}(1-\hat{p})} = \sqrt{(1/4)(3/4)} = \frac{\sqrt{3}}{4}$. So, the 95% confidence interval is $(\hat{\mu} - 2\hat{\sigma}/\sqrt{n}, \hat{\mu} + 2\hat{\sigma}/\sqrt{n})$, where $n = 100$ to get $(1/4 - \sqrt{3}/20, 1/4 + \sqrt{3}/20)$.

21. Every day, the number of people who are born is Poisson distributed with an average of 4900 people per day. We count how many people are born in a span of 100 days and let $\bar{X}$ denote the average number of people born per day. What is the probability $P(\bar{X} \leq 4895)$?

Solution: Since it is Poisson distributed, $\lambda = 4900$ and so the standard deviation is $\sqrt{\lambda} = \sqrt{4900} = 70$. By the Central Limit Theorem, the average number of people born in a span of 100 days is approximately normally distributed with expected value $\lambda = 4900$ and standard deviation $70/\sqrt{n} = 70/\sqrt{100} = 7$. So, $P(\bar{X} \leq 4895) = 1/2 - P(4895 \leq \bar{X} \leq 4900) = 1/2 - z(\frac{|4895-4900|}{7}) = 1/2 - z(5/7)$, where we look up $5/7$ in a z-score table.

3 Hypothesis Testing

22. **TRUE** False We never accept the null hypothesis, we just fail to reject it.

Solution: You can say we fail to reject or keep the null hypothesis. We don't accept or prove the null hypothesis.

23. True **FALSE** The Z test only works for random variable $X_1, \ldots, X_n$ normally distributed.
24. **TRUE** False The T test only works for random variable $X_1, \ldots, X_n$ normally distributed.
25. True **FALSE** The T test only works for $n < 30$.

Solution: It works for larger n but the t values become approximately equal to the Z values so we can use that table instead.

26. Which type of hypothesis test should you use for the following situations: (PMF Hypothesis Test/Z-Test/T-Test/χ^2 Goodness-of-fit/χ^2 Independence)

    1. You want to know if a die is fair so you roll it 100 times and count the number of $1, 2, \ldots, 6$s
    2. You want to know if a die is biased towards 5 so you roll it 100 times
    3. You want to know if a student's letter grade is related to whether they go to office hours
    4. You want to know if a coin is biased towards heads and flip it 10 times
    5. You want to know if flipping coins give you a binomial distribution so you ask 500 friends to flip a coin 5 times and count the number of friends who flipped $0, 1, \ldots, 5$ heads

6. You want to know if a coin is biased towards heads and flip it 100 times
7. You want to know if course evals are independent of the section so you count the number of $1, 2, \ldots, 7$s per section
8. You know heights are normally distributed and want to know if Berkeley students are taller than normal so you take 10 student heights

Solution:

1. χ^2 goodness of fit
2. Z test
3. χ^2 independence
4. PMF Hypothesis Test
5. χ^2 goodness of fit
6. Z test
7. χ^2 independence
8. t test

27. (True story) A woman claims that she can smell when someone has Parkinson's disease. She is given 10 people's shirts and correctly said whether the person had the disease in 9 of the 10 cases. Does she have this ability with $\alpha = 0.05$? (The 10th person who she said had Parkinson's actually developed it months later so she was really 10 for 10).

Solution: The null hypothesis would be that she guessed randomly and the probability that she was successful once is $p = \frac{1}{2}$. We want to calculate the probability that she did at least this well so $P(X \geq 9) = P(X = 9) + P(X = 10) = \binom{10}{9}\frac{1}{2^9}(1 - \frac{1}{2}) + \binom{10}{10}\frac{1}{2^{10}} = \frac{11}{2^{10}} = 0.01 < \alpha$. Therefore, we can reject the null hypothesis and say that she does have this ability.

28. When counting families with 2 children, I find that 83 of them have two girls, 102 of them have two boys, and 215 of them have one boy and one girl. Suppose that my null hypothesis is having a boy or girl is that I expect a $1 : 1 : 2$ ratio. Can we reject the null hypothesis with $\alpha = 0.05$?

Solution: This is a χ^2 test. The expected distribution is 100 families having 2 girls, 100 having two boys, and 200 having one boy and one girl. The χ^2 value is $\frac{(83-100)^2}{100} + \frac{(102-100)^2}{100} + \frac{(215-200)^2}{200} = 4.055$. We look up the critical value in a χ^2 table with $3-1 = 2$ degrees of freedom and $\alpha = 0.05$ to get 5.991 and compare it to 4.055. So we keep the null.

29. The height of 4 NBA players is 75 inches with a sample standard deviation of $s = 6$. Can we reject the null hypothesis that NBA players are the same height as the average person's height (which is 66 inches) with a two sided alternative hypothesis and $\alpha = 0.05$?

Solution: We want to calculate the probability that $P(X \geq 75)$ and the t statistic is $t(\frac{|75-66|}{6/\sqrt{4}}) = t(\frac{9}{83}) = t(3)$. We have 3 degrees of freedom and looking this up gives $t(3) = 0.029 > \alpha/2 = 0.025$. So we keep the null hypothesis that the heights of NBA players is no different.

30. Write the PDF for the t distribution when $\nu = 4$.

Solution:

$$f_4(x) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} = \frac{1}{\sqrt{4\pi}} \frac{\Gamma(5/2)}{\Gamma(2)} (1 + x^2/4)^{-5/2}$$

$$= \frac{15\sqrt{\pi}/8}{1 \cdot 2\sqrt{\pi}} (1 + x^2/4)^{-5/2} = \frac{15}{16} (1 + x^2/4)^{-5/2}.$$

31. You think that students who go to office hours will do better than those that don't. Suppose that exam scores are normally distributed with $\mu = 70$. You ask 9 students who go to office hours what their exam scores were and get a sample mean of $\bar{x} = 74$ and a sample standard deviation of $s = 6$. Can you say that students that go to office hours do better than other students with $\alpha = 0.05$?

Solution: We want to calculate the probability that $P(X \geq 75)$ and the t statistic is $t(\frac{|74-70|}{6/\sqrt{9}}) = t(\frac{4}{2}) = t(2)$. We have 8 degrees of freedom and looking this up gives $t(2) = 0.040 < \alpha$. So we reject the null hypothesis that the average for students who go to office hours should be the same.

32. Find $\chi^2(x)$ for $k = 1, 2, 4$.

Solution:

$$\chi^2_{k=1}(x) = \frac{1}{2^{1/2}\Gamma(1/2)} x^{1/2-1} e^{-x/2} = \frac{1}{\sqrt{2}\sqrt{\pi}} \frac{1}{\sqrt{x}} \frac{1}{\sqrt{e^x}} = \frac{1}{\sqrt{2\pi x e^x}}.$$

$$\chi^2_{k=2}(x) = \frac{1}{2^{2/2}\Gamma(2/2)} x^{2/2-1} e^{-x/2} = \frac{1}{2 \cdot 0!} x^0 e^{-x/2} = \frac{1}{2} e^{-x/2}.$$

$$\chi^2_{k=4}(x) = \frac{1}{2^{4/2}\Gamma(4/2)} x^{4/2-1} e^{-x/2} = \frac{1}{4 \cdot 1!} x^1 e^{-x/2} = \frac{1}{4} x e^{-x/2}.$$

33. Every year 25% of people contract the flu. This year, the NIH comes out with a vaccine and out of 1600 people, there are only 350 people who contract the disease. Was the vaccine successful with $\alpha = 1\%$?

Solution: The null hypothesis is the vaccine was not successful and hence the probability of getting the disease is $p = 0.25$. When we sample 1600 people, we expect a binomial distribution with $n = 1600$ and hence have a mean of $np = 400$ and standard deviation of $\sqrt{np(1-p)} = 10\sqrt{3}$. So the probability of getting at least an extreme case of 20 people is $\frac{1}{2} - z(|350 - 400|/(10\sqrt{3})) \approx 1/2 - z(2.89) = 0.5 - 0.4981 = 0.0019 < \alpha$. So, we reject the null hypothesis and say that the vaccine was successful.

34. You are wondering whether performing well in this course and gender are related and you get the following table. Are they related?

	Male	Female
Pass	315	485
Fail	85	115

Solution: There are a total of 800/1000 people who pass and 200/1000 who fail, and 400/1000 who are male and 600/1000 who are female. Thus, if they were independent, for instance we would expect that $\frac{800}{1000} \cdot \frac{600}{1000} = 48\%$ of people to be female and pass. We can fill out the expected table as follows:

	Male	Female
Pass	320	480
Fail	80	120

Now we can do the χ^2 test to get a statistic of

$$\frac{(315-320)^2}{320} + \frac{(485-480)^2}{480} + \frac{(85-80)^2}{80} + \frac{(115-120)^2}{120} = 0.651.$$

The critical value for 1 degree of freedom is 3.841 and $0.651 < 3.841$ so we cannot reject the null hypothesis.

35. In a skittle bag, you get 14 red skittles, 12 blue, 1 green, 10 yellow, and 13 orange skittles. Is it possible that the colors are evenly distributed with a significance level of $\alpha = 0.05$?

Solution: In 50 skittles, we expect to get 10 of each. Following the formula, our statistic is:

$$\frac{(14-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(1-10)^2}{10} + \frac{(10-10)^2}{10} + \frac{(13-10)^2}{10}$$
$$= \frac{16+4+81+0+9}{10} = 11.$$

There are 5 options so we have $5 - 1 = 4$ degrees of freedom. For 4 degrees of freedom and $\alpha = 0.05$, our critical value is 9.488. Since $11 > 9.488$, we reject the null hypothesis that the colors are evenly distributed.

4 MLE

36. True **FALSE** For MLE, setting $\frac{df}{dx} = 0$ will always give a minimum or maximum of f.

Solution: It will lead to critical points which won't always be mins or maxs (think x^3 at $x = 0$).

37. True **FALSE** The estimator you get from maximum likelihood estimation will be unbiased.

38. There is a bag with 10 balls colored red and blue. You pull out two balls (with replacement) and get BR. What is the maximum likelihood for the number of blue balls in bag?

Solution: Let n be the number of blue balls. Then the likelihood is given by $\frac{n}{10} \cdot \frac{10-n}{10}$. Taking the derivative and setting equal to 0 gives $n = 5$.

39. You have a coin that you think is biased. you flip it 3 times and get the sequence HTH. What is the maximum likelihood estimate for the probability of getting heads?

Solution: Let p be the probability of getting heads. Then the probability of getting HTH is $p^2(1-p)$. Taking the derivative and setting equal to zero gives $2p^2 - 3p^2 \implies p = \frac{2}{3}$.

40. You assume that the lifespan of lightbulbs are exponentially distributed (PDF is $\lambda e^{-\lambda t}$ for $t \geq 0$) and notice that your three light bulbs go out in 1, 2, and 3 years. What is the maximum likelihood estimator for λ?

Solution: We want to find the maximum likely λ given our sample of $1, 2, 3$. So, we want to maximize $L(\lambda|x_1, x_2, x_3)$ where $x_i = i$. By definition, we have that $L(\lambda|x_1, x_2, x_3) = P(x_1, x_2, x_3|\lambda) = P(x_1|\lambda)P(x_2|\lambda)P(x_3|\lambda)$ by independence. We calculate that as $\lambda^3 e^{-6\lambda}$. In order to find the maximum, we take the derivative and set it equal to 0 to get

$$3\lambda^2 e^{-6\lambda} - 6\lambda^3 e^{-6\lambda} = 0 \implies \lambda = 0, \frac{1}{2}.$$

The solution $\lambda = 0$ doesn't make sense and hence $\lambda = 1/2$.

41. I go to Kip's and want to figure out the total number of students n there. By looking, I see that I've taught 2 of the students there. I pick a student at random and it turns out to be one of the students I've taught. Then 8 more students come in and of those I have taught 7 of them. Now I again pick a random student and I haven't taught this second student picked. What is the most likely number n of total students at Kip's originally?

Solution: We want to find $L(n|pick)$ the likelihood of n given what I picked. This is just the probability. So originally, there are 2 students I've taught and $n-2$ that I haven't. The probability of picking a student I've taught is $\frac{2}{n}$. Then 8 more come in and I've taught a total of $2+7=9$ of them and not taught $n-2+1=n-1$ of them out of a total of $n+8$ students. So the probability of picking someone I didn't teach is $\frac{n-1}{n+8}$. So the likelihood function is

$$L(n) = \frac{2}{n} \cdot \frac{n-1}{n+8} = \frac{2n-2}{n^2+8n}.$$

We take the derivative and set it equal to 0. This gives

$$\frac{(n^2+8n)\cdot 2 - (2n-2)(2n+8)}{(n^2+8n)^2} = \frac{-2(n-4)(n+2)}{(n^2+8n)^2} = 0.$$

So we get $n = 4$ or $n = -2$. But $n = -2$ which doesn't make sense. Then note that f' is positive before and negative after so this is a local and global max. So the maximum likelihood estimate is $\hat{n} = 4$.

5 Miscellaneous

42. **TRUE** False The line of best fit is the line that minimizes the least square error.

43. Suppose that three people randomly pick a hat. What is the expected value of the number of people who choose their hat? (with proof). What is the variance? Now do the same with n people.

Solution: Use $X = X_1 + X_2 + X_3$ where X_i is Bernoulli on whether person i gets their hat back. Then we calculate $E[X_i] = \frac{1}{3}$ and $E[X_i X_j] = \frac{1}{6}$ for $i \neq j$. This gives that expected value and variance are both 1.

44. Stirling's approximation tells us that $n! \approx \frac{n^n\sqrt{2n\pi}}{e^n}$. Use this to show that $\binom{2n}{n} \approx \frac{2^{2n}}{\sqrt{n\pi}}$.

Solution:

$$\begin{aligned}
\binom{2n}{n} &= \frac{(2n)!}{n!n!} \\
&\approx \frac{(2n)^{2n}\sqrt{2(2n)\pi}e^{-2n}}{n^n\sqrt{2n\pi}e^{-n}n^n\sqrt{2n\pi}e^{-n}} \\
&= \frac{2^{2n}n^{2n}\sqrt{4n\pi}e^{-2n}}{n^{2n}\sqrt{2n\pi}e^{-2n}\sqrt{2n\pi}} \\
&= \frac{2^{2n}\sqrt{2}}{\sqrt{2n\pi}} \\
&= \frac{2^{2n}}{\sqrt{n\pi}}
\end{aligned}$$

45. Prove that $\Gamma(x+1) = x\Gamma(x)$ for all $x > 0$.

Solution: We integrate by parts with $u = t^x$ and $dv = e^{-t}dt$ so $du = xt^{x-1}$ and $v = -e^{-t}$ to get

$$\begin{aligned}
\Gamma(x+1) &= \int_0^\infty t^x e^{-t} dt \\
&= t^x(-e^{-t})|_0^\infty + \int_0^\infty xt^{x-1}e^{-t}dt \\
&= 0 - 0 + x\int_0^\infty t^{x-1}e^{-t}dt = x\Gamma(x).
\end{aligned}$$

46. Use induction to prove that $\Gamma(n) = (n-1)!$ for all $n \geq 1$.

Solution: We prove the base case of $n = 1$. We get $\Gamma(1) = \int_0^\infty t^0 e^{-t} dt = \int_0^\infty e^{-t} dt = 1 = 0!$. Now assume the inductive hypothesis for some $n \geq 1$. Then, we have that

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)! = n! = (n+1-1)!.$$

Thus by mathematical induction, we have shown the result for all $n \geq 1$.

47. Use induction to prove that $E[\chi^2_{k=2n}(x)] = 2n$ for all $n \geq 1$.

Solution: Show the base case then use integration by parts with $u = x^n$. Come to office hours for details.

48. What is the definition of variance? Prove that $Var(X) = E[X^2] - E[X]^2$.

Solution: By definition, $Var(X) = E[(X - E[X])^2]$.

$$E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] = E[X^2] - 2E[XE[X]] + E[E[X]^2].$$

Now $E[X]$ is a constant and $E[c] = c$ so we can get rid of expected values to get

$$= E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2.$$

49. What is the definition of covariance? Prove that $Cov(X, Y) = E[XY] - E[X]E[Y]$.

Solution:

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY - XE[Y] - YE[X] + E[X]E[Y]]$$

$$= E[XY] - E[XE[Y]] - E[YE[X]] + E[E[X]E[Y]].$$

Then since $E[X], E[Y]$ are constants and $E[cX] = cE[X]$, we can take them out to get

$$= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] = E[XY] - E[X]E[Y].$$

50. The formulas for the slope and y intercept of the line of best fit come from MLE. Suppose that error is normally distributed. This means that if we predict $y = ax_i + b$, then the probability of actually getting y_i follows the PDF

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-(y_i - y)^2/2\sigma^2} = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y_i - (ax_i + b))^2/2\sigma^2}.$$

Use MLE to show that $\hat{b} = \bar{y} - a\bar{x}$.

Solution: We calculate $L(\theta|(x_1, y_1), \ldots, (x_n, y_n))$ below

$$\begin{aligned} L(\theta|(x_1, y_1), \ldots, (x_n, y_n)) &= P((x_1, y_1), \ldots, (x_n, y_n)|b = \theta) \\ &= \prod_i \frac{1}{\sigma\sqrt{2\pi}} e^{-(y_i - (ax_i + \theta))^2/2\sigma^2} \\ &= \frac{1}{\sigma^n \sqrt{2\pi}^n} e^{-\sum (y_i - (ax_i + \theta))^2/2\sigma^2} \end{aligned}$$

Now taking the log gets rid of the exponent and taking the derivative and setting it equal to 0 gives

$$\begin{aligned} 0 &= -\sum \frac{\partial}{\partial\theta} \frac{(y_i - ax_i - \theta)^2}{2\sigma^2} \\ &= \sum \frac{2(y_i - ax_i - \theta)}{2\sigma^2} \end{aligned}$$

So $\sum(y_i - ax_i - \theta) = \sum(y_i - ax_i) - n\theta = 0$ and so

$$\theta = \hat{b} = \frac{1}{n} \sum (y_i - ax_i) = \bar{y} - a\bar{x}.$$

51. Now with $b = \bar{y} - a\bar{x}$, do MLE to show that $\hat{a} = r\frac{\sigma_y}{\sigma_x}$ the formula that we use for a.

Solution: We calculate $L(\theta|(x_1, y_1), \ldots, (x_n, y_n))$ below

$$\begin{aligned} L(\theta|(x_1, y_1), \ldots, (x_n, y_n)) &= P((x_1, y_1), \ldots, (x_n, y_n)|a = \theta) \\ &= \prod_i \frac{1}{\sigma\sqrt{2\pi}} e^{-(y_i - (\theta x_i + (\bar{y} - \theta\bar{x})))^2/2\sigma^2} \\ &= \frac{1}{\sigma^n \sqrt{2\pi}^n} e^{-\sum ((y_i - \bar{y}) + \theta(\bar{x} - x_i))^2/2\sigma^2} \end{aligned}$$

Now taking the log gets rid of the exponent and taking the derivative and setting it equal to 0 gives

$$\begin{aligned} 0 &= -\sum \frac{\partial}{\partial\theta} \frac{((y_i - \bar{y}) + \theta(\bar{x} - x_i))^2}{2\sigma^2} \\ &= \sum \frac{2(\bar{x} - x_i)((y_i - \bar{y}) + \theta(\bar{x} - x_i))}{2\sigma^2} \end{aligned}$$

Simplifying gets the result.