NYCCT MAT 2572 Halleck Spring 2016 Practice exam 3 solutions

You may submit a formula sheet (1 sheet, 2 pages, hand-written), worth up to 5% extra.

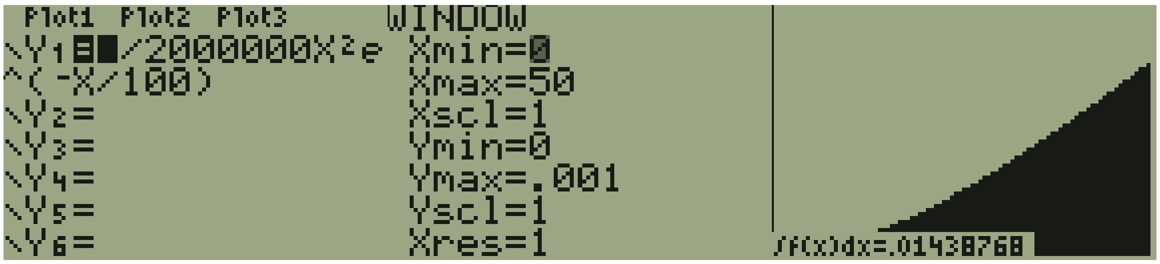
Distribution	$f_Y(y)$	$M_X(t)$
Binomial B(n, p)	$p(k) = \binom{n}{k} p^k q^{n-k}$	$(q + pe^t)^n$
Geometric (counting failures only)	$p(k) = pq^k$	$\frac{p}{1 - qe^t}$
Negative Binomial (counting failures)	$p(k) = \binom{r+k-1}{r-1} q^k p^r$	$\left(\frac{p}{1-qe^t}\right)^r = \frac{p^r}{(1-qe^t)^r}$
Poisson	$p(k) = \lambda^k e^{-\lambda}/k!$	$e^{\lambda(e^t - 1)}$
Uniform	$f_Y(y) = \frac{1}{(b-a)}; a \le y \le b$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$
Uniform discrete	$p(k) = \frac{1}{(b-a+1)}; a \le k \le b$	$\frac{e^{t(b+1)} - e^{ta}}{(e^t - 1)(b-a+1)}$
Normal	$f_Y(y) = \frac{1}{\sigma\sqrt{2p}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$	$e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
Exponential	$f_Y(y) = \lambda e^{-\lambda y}$	$(1 - t\lambda^{-1})^{-1}$
Gamma	$f_Y(y) = \lambda^r/(r-1)!\, y^{r-1} e^{-\lambda y}$	$(1 - t\lambda^{-1})^{-r}$

Part I: You will be given 2 out of the following 3 types of problems.

1. (25 pts) Recently married, a young couple plans to continue having children until they have their first girl. Suppose the probability that a child is a girl is 1/2, the outcome of each birth is an independent event, and the birth at which the first girl appears has a geometric distribution. What is the couple's expected family size? Is the geometric pdf a reasonable model here? Discuss.
According to the above table, the MGF for counting just the failures (which in this case are the boy children) is $\frac{p}{1-qe^t} = p(1 - qe^t)^{-1}$. Taking the derivative gives $pqe^t(1 - qe^t)^{-2}$ and setting t=0 gives q/p. However, our distribution Y is a count of all the family members so need to add 1 for the girl and 2 for the parents; hence E(Y)=q/p +3=(1-p)/p+3=1/p+2=4. Thus, the couple's expected family size is 4. The model does not take into account that a woman can have a limited number of children within her lifetime, perhaps 20 or 30, assuming a single child is born each time. The world record is 69, but that was from a woman who had multiple children births many times.

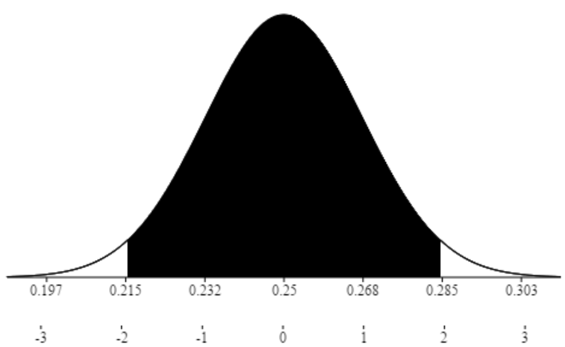
2. (25 pts) A machine has a 1% probability of producing a defective item. Each day, the machine is run until a defective item is produced and then it undergoes an extensive adjustment which requires the rest of the working day. What is the average number of items that will be produced in that work week? In a given 5-day workweek, what is the probability that a machine will produce 500 or more usable items? (Negative binomial)
Identifying the 3 parameters that we need, r=5, p=.01, q=.99, our MGF is $\left(\frac{p}{1-qe^t}\right)^r = p^r(1 - qe^t)^{-r}$ and taking the derivative gives: $rp^r qe^t(1 - qe^t)^{-r-1}$. Setting t=0, gives rq/p which in our case is 5*.99/.01=495. However, this is just the number of usable items produced. We also need to add the defectives to get a total of 500. To find the chance that 500 or more USABLE are produced, we use software. Our inputs are 500 failures (usables) and r=5 successes (defectives). For instance with Excel, the command is $P(Y \ge 500) = 1 - P(Y \le 499) =$ 1-NEGBINOM.DIST(499,5,0.01,TRUE)=43%. We know that the average is 495 failures but even $P(Y \ge 495) = 1 - P(Y \le 494) =$ 1-NEGBINOM.DIST(494,5,0.01,TRUE)=44% is less than 50%. The reason is that distribution is skewed to the right so average is larger than the median (which is ~462, see Excel file).

3. (25 pts) The next generation of space shuttle will include three fuel pumps —one active, the other 2 in reserve. If the primary pump malfunctions, a second is automatically brought on line. Suppose a typical mission is expected to require that fuel be pumped for at most 50 hours. According to the manufacturer's specifications, pumps are expected to fail once every 100 hours. If the pumps were allowed to go until all 3 failed, on average how many hours will that be? What are the chances that such a fuel pump system would not remain functioning for the full 50 hours that might be required?
The MGF is $\left(1 - t\lambda^{-1}\right)^{-r}$ and taking the derivative, we get $\frac{r}{\lambda}\left(1 - t\lambda^{-1}\right)^{-r-1}$ and setting t=0, we get $\frac{r}{\lambda} = \frac{3}{1/100} = 300$
To calculate that the system will remain functioning for 50 hours, we use the pdf
$$\frac{1}{2*100^3}\int_0^{50} y^2 e^{-y/100}dy = 0.0144\sim 1.4\%$$



Part II You will be given 2 out of the following 3 types of problems. No picture no credit for this section.

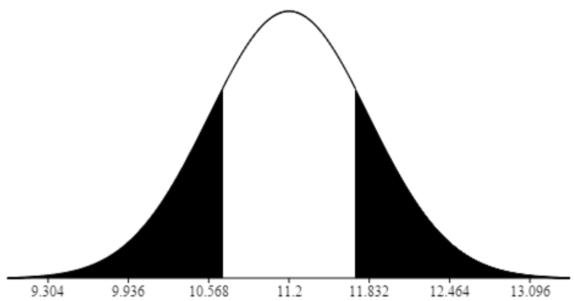
4. (25 pts) A poll taken on Tuesday 11/24/15
http://www.quinnipiac.edu/news-and-events/quinnipiac-university-poll/iowa/release-detail?ReleaseID=2305
has Trump leading the Iowa Caucus polls 25% to Rubio’s 23%. However, the lead is within the claimed margin of error of $\pm$ 4%. Focus just on Trump’s numbers.
   a. Find the standard deviation for whether one voter is a supporter of Trump or not.
   This is a binomial distribution (coin flipping) with just one trial (flip).
   σ= √(pq)= √(.25*.75)=.43
   b. Given that the sample size was 600, find the standard error (the standard deviation for the sample mean). **Std err=σ/√n=.43/√600=.0177** Draw the appropriate normal curve and label the horizontal axis with both Xbar and Z labels.



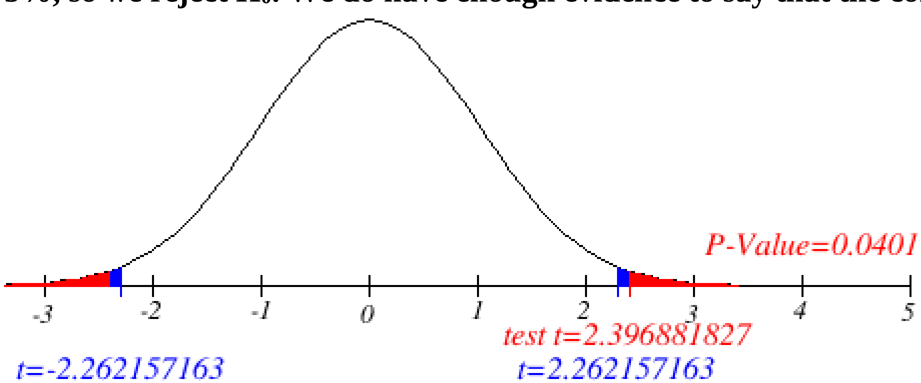
   c. The margin of error is approximately $\pm$ twice the standard error. Find it. Is it roughly what the polltakers claim? **Margin of error = ±2(.0177)= ±3.5% which if we round up is the 4% provided by the polltakers.**
   d. By what factor would the sample size have had to be increased (assuming that results will stay the same) for us to conclude that Trump was definitely leading. The difference between the candidates is 2%. **To definitively say that Trump was in the lead, we need to have quartered**

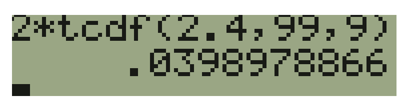
the margin of error which corresponds to quartering the standard error. To quarter the standard error, we need to increase the sampling size by a factor of 16.

5. (25 pts) A previous sample of fish in Lake Michigan indicated that the mean polychlorinated biphenyl (PCB) concentration per fish was 11.2 parts per million with a standard deviation of 2 parts per million. Suppose a new random sample of 10 fish has the following concentrations:
11.5, 12.0, 11.6, 11.8, 10.4, 10.8, 12.2, 11.9, 12.4, 12.6
Test the hypothesis that the mean PCB concentration has also remained.
Don't forget to make a histogram to verify that the data are from a normal distribution.
NOTE: this is a 2-tailed test.
H_0: $\mu = 11.2$, H_1: $\mu \neq 11.2$ Sample mean is 11.72. Std err=$\sigma/\sqrt{n}=2/\sqrt{10}=.632$
So the z value for the sample mean is z=(11.72-11.2)/.632=.822
and the pvalue=2*(1-norm.s.dist(.822,true))=.411~41%
This is way above 5%, so we do not reject H_0. We do not have enough evidence to say that the concentration has changed.



5b.(25 pts) A previous sample of fish in Lake Michigan indicated that the mean polychlorinated biphenyl (PCB) concentration per fish was 11.2 parts per million ~~with a standard deviation of 2 parts per million~~. Suppose a new random sample of 10 fish has the following concentrations:
11.5, 12.0, 11.6, 11.8, 10.4, 10.8, 12.2, 11.9, 12.4, 12.6
Test the hypothesis that the mean PCB concentration has also remained.
Don't forget to make a histogram to verify that the data are from a normal distribution.
NOTE: this is a 2-tailed test.
H_0: $\mu = 11.2$, H_1: $\mu \neq 11.2$ Sample mean is 11.72. Std err=$s/\sqrt{n}=.686/\sqrt{10}=.217$
So the t value for the sample mean is z=(11.72-11.2)/.217=2.40
and the pvalue=2*(1-tdist(2.4,true))=.040~4%(using Excel). For calc see below. This is below 5%, so we reject H_0. We do have enough evidence to say that the concentration has changed.





6. (25 pts) A farmer claims to be able to produce larger tomatoes. To test this claim, a tomato variety that has a mean diameter size of 8.2 centimeters with a standard deviation of 2.4 centimeters is used. A sample size of 36 tomatoes will be used to test claim. If the actual mean is 9.1 centimeters, calculate the power of the test to show that the mean size is indeed larger. Assume that the population standard deviation remains equal to 2.4.
NOTE: this is a 1-tailed test.

H_0: μ = 8.2, H_1: > 8.2

Std err= $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.4}{\sqrt{36}} = 0.4$

a. **Work with the normal distribution for the sample mean centered around 8.2, which is the average size for this variety of tomato.**
   i. **To calculate the power, we need the critical value:**
   $$z_{.05} = \text{norm.s.inv}(.95) = 1.64$$
   ii. **Destandardize: $\sigma_{\overline{X}} * z_{.05} + \mu = 0.4 * 1.64 + 8.2 = 8.86$**

b. **Switch to the sample mean distribution around 9.1 (the actual mean for this farmer). The chance that the sample mean for our farmer's tomatoes is bigger than the critical value will be our power= $1 - \beta$.**
   i. **In the sample mean distribution for our farmer's tomatoes, the standardized version of the critical value is $\frac{8.86-9.1}{.4} = -.605$**
   ii. **hence**
   $$\text{power} = 1 - \beta = 1 - \text{norm.s.dist}(-.605) = 1 - .373 = .727 \approx 73\%$$

Thus, 73% of the time we will correctly say that the farmer's tomatoes are on average bigger than the varietal average in general.