Math 50: Final

180 minutes, 140 points. No algebra-capable calculators. Try to use your calculator only at the end of your calculation, and show working/reasoning. Please *do* look up z, t, χ^2 values for full credit. Attempt all questions, but in the order most comfortable for you, and heed the number of points available. Later question parts may be independent of earlier ones, so skip over one you can't do.

1. [13 points] It was found that 35 out of 300 'famous' people have the star sign Sagittarius.

   (a) Test the null hypothesis that Sagittariuses are no more likely to become famous than any other star sign (assume each sign is exactly 1/12 of the year), against the hypothesis that they are more prone to fame and fortune.

   (b) Give this same data, what would a Bayesian, with no prior bias towards any particular value of the proportion p of famous Sagittariuses, conclude about p? (That is, what is the normalized posterior on p?)

2. [15 points] Consider the joint pdf $f_{X,Y}(x, y) = 6y$, for $0 < x < 1$ and $0 < y < 1-x$. [Hint: double-check your domain].

   (a) Find the marginal pdf $f_X(x)$.

   (b) Find the conditional pdf $f_{Y|x}(y)$.

   (c) Find the covariance $\text{Cov}(X, Y)$.

3. [14 points]

(a) How many ways are there of choosing a team of 3 and a team of 5 out of a class of 8?

[2-point Bonus: if the teams are both of size 4 and not distinguishable, *e.g.* not labelled A or B, how many ways are there now?]

(b) A shipment of 24 eggs has 6 bad ones. You test the shipment by picking 3 at random and accepting it if at most one is bad. What is the chance that you accept the shipment?

(c) You have 3 drawers of socks: A contains two white, B contains a white and a black, and C contains two black. However, you can't remember which drawer is which. Suppose you open a random drawer and pick out a random sock. That sock is white. What are the probabilities that the drawer you opened are A, B, or C? Please explain, using probability notation, and stating what rule of probability you used.

4. [18 points] Rectangular mint candies of variable size are produced by a machine. Let X and Y be random variables giving their width and length.

   (a) If X has uniform pdf in the interval $[0, 1]$ and, independently, Y has uniform pdf in the interval $[0, 2]$, find the probability that $Y \geq X$.

(b) Given the same pdf as above, find the pdf of the candy *area* $A = XY$.

(c) The edge of each candy is lined by chocolate, with length given by the perimeter $P = 2(X + Y)$. Given the same pdf as above, find the pdf of the perimeter. [Hint: find pdf of $X + Y$ first, taking care to consider its various domains].

(d) Finally, assume the machine is adjusted so that the candies are each *square*, so there is only one random variable, X, and it has uniform pdf in $[0, 1]$. Find the new pdf of the area $A = X^2$.

5. [25 points] Consider the model pdf $f_Y(y;\theta) = (1/\theta)e^{-y/\theta}$ for $y \geq 0$. Data $\{y_i\}$, $i = 1, \ldots, n$ are collected, from which we wish to estimate θ.

(a) Derive the Maximum Likelihood estimator $\hat{\theta}$. Please show your working.

(b) Compute the variance of this estimator, $\mathrm{Var}(\hat{\theta})$, assuming the data does in fact come from the model pdf with parameter value θ.

(c) Compare this to the Cramér-Rao lower bound. Is then $\hat{\theta}$ an *efficient* estimator?

(d) Prove that the estimator is *consistent*. [Hint: make sure you demonstrate everything you need to].

(e) With the same assumption as before, compute the full normalized pdf of the estimator, $f_{\hat{\theta}}(u)$. [Hint: go back to the form of $\hat{\theta}$].

(f) What family is the conjugate prior, and why?

6. [25 points] Heights of a random sample of four US males were 70, 72, 62, 68 inches.

   (a) US females have an approximately normal height distribution with mean 64, variance 6. Assuming the male variance is the same as for females, use the data to test the hypothesis that μ for males is *greater* than that of females, at the 95% confidence level.

   (b) What is your p-value for the above test?

   (c) Instead assume the male variance σ^2 is unknown, and compute the *95% confidence interval* on the mean μ for US males.

(d) Assuming the underlying pdf is normal, and μ is unknown, compute a *80% confidence interval* on the variance σ^2.

(e) Can you reject the null hypothesis that $\sigma^2 = 6$ (that of US females), at the 80% confidence level, compared against the hypothesis that $\sigma^2 \neq 6$?

7. [15 points] Responses to the survey question, "Do you like cheese?" gave 147 out of 210 responding "Yes". (The rest responded, conveniently enough, "No").

   (a) Give a 95% confidence interval on the underlying proportion p of the population that 'like' cheese.

   (b) What assumption(s) is/are needed to justify this last conclusion?

   (c) After a brutal nationwide marketing campaign by the Cheese Board of America (no pun intended), a new survey finds 118 out of 150 responding "Yes". Test the hypothesis that the proportion has increased, against the null hypothesis that it remained the same, at the 95% confidence level.

8. [15 points] A computer LCD screen contains 1 million (that is, 10^6) pixels, each of which has an independent probability p of being 'dead' due to manufacturing defects. LCD panels are considered acceptable if they have at most 3 dead pixels.

(a) Say $p = 10^{-6}$. What is the average percentage of LCD panels produced that are *unacceptable*?

(b) What is the largest p can be if the factory must produce on average at least 50% of panels which have *no* dead pixels?

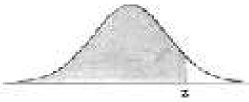
(c) Quality Control examines 100 panels and counts 200 dead pixels in total. Use this data, and possibly additional assumptions, to construct a *95% confidence interval* on p.

Useful formulae and pdfs:

$$
\begin{aligned}
f_{Y_i'}(y) &= \frac{n!}{(i-1)!(n-i)!} F_Y(y)^{i-1}[1-F_Y(y)]^{n-i} f_Y(y) \\
f_W(w) &= \int f_X(x) f_Y(w-x)dx \qquad \text{for} \quad W = X+Y \\
f_W(w) &= \int \frac{1}{|x|} f_X(w/x) f_Y(x)dx \qquad \text{for} \quad W = XY \\
f_W(w) &= \int |x| f_X(x) f_Y(wx)dx \qquad \text{for} \quad W = Y/X \\
\text{poisson } p_X(k;\lambda) &= e^{-\lambda}\frac{\lambda^k}{k!}, \text{ for } k = 0,1,2,\ldots, \quad \lambda \geq 0, \quad E(X) = \text{Var}(X) = \lambda \\
\text{gamma } f_Y(y;r,\lambda) &= \frac{\lambda^r}{\Gamma(r)} y^{r-1} e^{-\lambda y}, \text{ for } y \geq 0, \qquad E(Y) = \frac{r}{\lambda}, \quad \text{Var}(Y) = \frac{r}{\lambda^2} \\
\text{beta } f_Y(y;r,s) &= \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} y^{r-1}(1-y)^{s-1} \text{ for } 0 \leq y \leq 1, \quad E(Y) = \frac{r}{r+s} \\
\text{normal } f_Y(y;\mu,\sigma) &= \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad E(Y) = \mu, \quad \text{Var}(Y) = \sigma^2 \\
\text{negative binomial } p_X(k;r,p) &= \binom{k-1}{r-1} p^k (1-p)^{k-r}, \text{ for } k = r, r+1, \ldots, \quad E(X) = \frac{r}{p}, \quad \text{Var}(X) = \frac{r(1-p)}{p^2}
\end{aligned}
$$

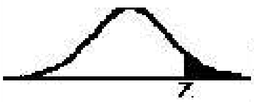
CDF of standard normal follows on next page, and 'far-right tail probabilities' which are $1 - F_Z(z)$ for large z up to 9.5.

Note you can get $F_Z(z)$ for $z < 0$ via $1 - F_Z(-z)$.



Probability Content
from -oo to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990



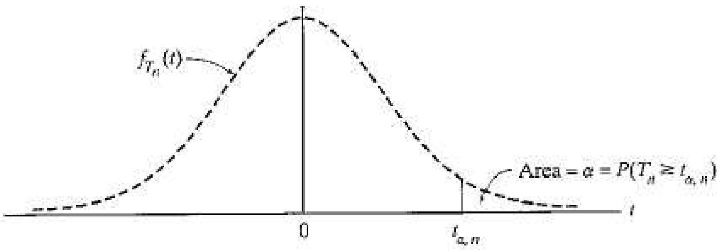
Far Right
Tail Probabilities

Z	P{Z to oo}	Z	P{Z to oo}	Z	P{Z to oo}	Z	P{Z to oo}
2.0	0.02275	3.0	0.001350	4.0	0.00003167	5.0	2.867 E-7
2.1	0.01786	3.1	0.0009676	4.1	0.00002066	5.5	1.899 E-8
2.2	0.01390	3.2	0.0006871	4.2	0.00001335	6.0	9.866 E-10
2.3	0.01072	3.3	0.0004834	4.3	0.00000854	6.5	4.016 E-11
2.4	0.00820	3.4	0.0003369	4.4	0.000005413	7.0	1.280 E-12
2.5	0.00621	3.5	0.0002326	4.5	0.000003398	7.5	3.191 E-14
2.6	0.004661	3.6	0.0001591	4.6	0.000002112	8.0	6.221 E-16
2.7	0.003467	3.7	0.0001078	4.7	0.000001300	8.5	9.480 E-18
2.8	0.002555	3.8	0.00007235	4.8	7.933 E-7	9.0	1.129 E-19
2.9	0.001866	3.9	0.00004810	4.9	4.792 E-7	9.5	1.049 E-21

t values corresponding to upper tail probabilities:

	α						
df	.20	.15	.10	.05	.025	.01	.005
1	1.376	1.963	3.078	6.3138	12.706	31.821	63.657
2	1.061	1.386	1.886	2.9200	4.3027	6.965	9.9248
3	0.978	1.250	1.638	2.3534	3.1825	4.541	5.8409
4	0.941	1.190	1.533	2.1318	2.7764	3.747	4.6041
5	0.920	1.156	1.476	2.0150	2.5706	3.365	4.0321
6	0.906	1.134	1.440	1.9432	2.4469	3.143	3.7074
⋮			⋮				
30	0.854	1.055	1.310	1.6973	2.0423	2.457	2.7500
∞	0.84	1.04	1.28	1.64	1.96	2.33	2.58

FIGURE 7.4.1



χ^2 values corresponding to lower and upper tail probabilities:

Figure 7.5.2 shows the top portion of the chi square table that appears in Appendix A.3. Successive rows refer to different chi square distributions (each having a different number of degrees of freedom). The column headings denote the areas *to the left* of the numbers listed in the body of the table.

We will use the symbol $\chi^2_{p,n}$ to denote the number along the horizontal axis that cuts off to its left an area of p under the chi square distribution with n degrees of freedom.

	p							
df	.01	.025	.05	.10	.90	.95	.975	.99
1	0.000157	0.000982	0.00393	0.0158	2.706	3.841	5.024	6.635
2	0.0201	0.0506	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.336	26.217