Assignment 8 (two pages plus two tables)

Statistics 24400 (Autumn 2024)

Due on Gradescope, Thursday, December 5 by 9 am

Reminder

Final exam: 12:30-2:30 pm, Thursday, December 12, E133. More details will be provided in class and on Ed Discussion.
Week 10 TA office hours: 4:30-5:30 pm Monday Dec 2 in Stuart 101; 4:30-5:30 pm Wednesday Dec 4 in Jones 303.

Problem assignments (related sections in the text: 8.5-8.6, 9.1-9.3)

1. (*Bayesian inference*) Suppose $X_1, \cdots, X_n$ are *i.i.d.* $\sim$ *Geometric*(Θ) with probability mass function (PMF)

$$f_X(x \mid \Theta) = \mathbb{P}(X_i = x \mid \Theta) = (1-\Theta)^{x-1}\Theta, \qquad x = 1, 2, \cdots.$$

Assume the prior distribution for Θ is the Beta distribution with mean $\frac{\alpha}{\alpha+\beta}$ and density

$$f_\Theta(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, \qquad \text{for } 0 \le \theta \le 1.$$

   (a) Find $f_{X|\Theta}(x, \cdots, x_n \mid \theta)$, the likelihood function (i.e. the joint PMF) of $X_1, \cdots, X_n$ given θ.

   (b) Find $f(x, \cdots, x_n, \theta) = f_{X|\Theta}(x, \cdots, x_n \mid \theta) f_\Theta(\theta)$, the joint distribution of $\{X_1, \cdots, X_n, \Theta\}$.

   (c) Find the posterior distribution of Θ given the data $X_1, \cdots, X_n$, write out its density function.

   (d) Find the posterior mean of Θ given the data $X_1, \cdots, X_n$.

2. (*Uniform MLE*) Suppose that $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Uniform}[-\theta, \theta]$, where $\theta \in (0, \infty)$ is an unknown parameter.

   (a) Calculate the family of densities, $f(x \mid \theta)$, and the likelihood of the data, $\prod_i f(X_i \mid \theta)$.

   (b) Find the MLE $\hat{\theta}$.

   (c) Fix some $\epsilon \in (0, 1)$. Calculate $\mathbb{P}(\hat{\theta} < (1-\epsilon)\theta)$, the probability that $\hat{\theta}$ is a substantialunderestimate of θ.

   (d) Fix some $\epsilon \in (0, 1)$. Calculate $\mathbb{P}(\hat{\theta} > (1+\epsilon)\theta)$, the probability that $\hat{\theta}$ is a substantialoverestimate of θ.

3. (*Normal variance LR Test*) Suppose that we have data $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(0, \theta)$ (here θ is the variance), and we are testing $H_0 : \theta = 1$ against $H_1 : \theta = 2$.

   (a) Compute LR, the likelihood ratio statistic for this test, simplified into a clean form.

   (b) Recall that we reject H_0 if $LR \le c$ for some threshold c.
   Compute the Type I error of the LR test at sample size $n = 40$, if we use the threshold $c = 1$.

   (c) Compute the Type II error of the LR test at sample size $n = 40$, if we use the threshold $c = 1$.

   (d) Suppose you observe data $X_1, \ldots, X_n$ for sample size $n = 40$ with these summary statistics: sample mean $\bar{X} = 0.8$, sample variance $S^2 = 0.9$. What is the value of the likelihood ratio test statistic for this data?

4. (*LR Test*) You observe data $X_1, \ldots, X_n$ lying in the interval $[0, 1]$. You want to test $H_0 : X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Uniform}[0, 1]$ against $H_1 : X_1, \ldots, X_n$ drawn i.i.d. with density $f(x) = 6x(1-x)$ on $x \in [0, 1]$.

   (a) What is the likelihood ratio?

   (b) Suppose that $n = 1$. Suppose we set a threshold $c = 1$ for the likelihood ratio. Let X be the one data point observed. What is the range of X values such that you would choose H_0, and what is the range of X values such that you would choose H_1?

   (c) Continuing for the case $n = 1$, what is the Type I error of this likelihood ratio test?

   (d) Continuing for the case $n = 1$, what is the Power of this likelihood ratio test?

5. (*MLE for a multinomial distribution*) Suppose we observe the following counts in course registration:

	Undergrads	Master's students	PhD students	total
Number	U	M	P	n

Let's think of the total class size n as fixed in advance, and then when registration is opened, each student registration is a random event (a multinomial random variable), which has some probability of being an undergrad or a masters or a PhD student. If the three probabilities are p_u, p_m, p_p then the likelihood of observing $U = u$ and $M = m$ and $P = p$ as our counts (for some nonnegative integers u, m, p that add up to n) is given by

$$\frac{n!}{u!\,m!\,p!}\, p_1^u\, p_2^m\, p_3^p$$

(This is a generalization of the binomial distribution, which has two categories—"successes" and "failures"—with probabilities p and $1-p$. The multinomial can have any number ≥ 2 of categories. Note that the probabilities must add up to 1, i.e., $p_1 + p_2 + p_3 = 1$ in this example above.)

Now suppose we believe that we should expect twice as many master's students as PhD's. Our hypothesis says that the three probabilities that parametrize our multinomial distribution, can be written as $1 - 3p$, $2p$, and p for some parameter p. Calculate the likelihood of p in terms of the observed U, M, P and the total n, then find the MLE $\hat{p}$.

6. (*Testing the binomial hypothesis*) In this problem you will learn to test hypotheses about multinomial data—that is, data where each observation falls into one of several categories. In Rice's textbook, this material is primarily in section 9.5, and the week 10 lectures will cover this material, but to do this problem all the information you need is described in the problem itself. Suppose that 50 students each have 4 tries to throw a dart at a target.

Here are the results:

# successes	0	1	2	3	4
# students	9	12	11	14	4

This data is multinomial—for each student, his result falls into one of these 5 categories.

Our goal is to answer the following question: is the distribution of this data characterized by a Binomial distribution with some common parameter p? That is, is it true that for each student, their 4 trials are independent and all have the same chance of success?

One reason to believe that this hypothesis is not true, would be if we think that different students have substantially different skill levels for throwing darts. For example, if we believe that most students are all either experts at darts or very poor at throwing darts, then we'd expect to see lots of 4's (for the experts) and lots of 0's and 1's (for those very bad at darts) and relatively few 2's and 3's, while for a Binomial distribution this could never be the case.

(a) First, let X_i be the number of successes for the ith student, for $i = 1, \ldots, 50$.
What is the likelihood for the observed data $X_1, \ldots, X_{50}$ using the parameter p?

(b) Now calculate the MLE $\hat{p}$ as a function of $X_1, \ldots, X_{50}$.

(c) Next, calculate the MLE for the actual observed data.

(d) Next, given this MLE, how many students would you expect to fall into each of the five categories? That is, for each category, calculate $\mathbb{E}(\text{\# of students in this category})$ if in fact the true p were equal to your MLE $\hat{p}$:

# successes	0	1	2	3	4
$\mathbb{E}$(# students)	??	??	??	??	??

Note that these expected values may be non-integers.

(e) Now, we'll compute a test statistic. If the numbers you observed (9, 12, etc) are very far from the expected values calculated in the previous part, then that suggests that the data might not be Binomial. However, the size of the discrepancy should be relative to the expected number (i.e. if you expect 21 and get 26, that's not as much of a discrepancy as if instead you expect 1 and get 6.) Our test statistic is therefore defined as

$$\sum_{\text{every category in the table}} \frac{(\text{observed value} - \text{expected value})^2}{\text{expected value}}.$$

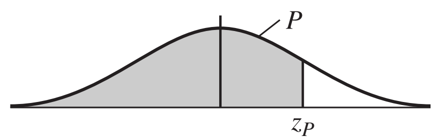
Calculate the value of the test statistic for the data observed in this problem.

(f) The statistic you calculated has a null distribution that is approximately a χ^2 distribution, where the number of degrees of freedom (d.f.) is equal to

(number of categories in the table) $-$ 1 $-$ (number of parameters fitted in your model).

Calculate this d.f.—now we know the (approximate) null distribution of the test statistic. Finally, use the attached χ^2 table from the textbook to get a p-value.

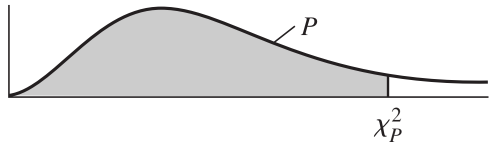
TABLE 2 Cumulative Normal Distribution—Values of P Corresponding to z_p for the Normal Curve



z is the standard normal variable. The value of P for $-z_p$ equals 1 minus the value of P for $+z_p$; for example, the P for –1.62 equals 1 – .9474 = .0526.

z_p	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.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
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

TABLE 3 Percentiles of the χ^2 Distribution—Values of χ^2_P Corresponding to P



df	$\chi^2_{.005}$	$\chi^2_{.01}$	$\chi^2_{.025}$	$\chi^2_{.05}$	$\chi^2_{.10}$	$\chi^2_{.90}$	$\chi^2_{.95}$	$\chi^2_{.975}$	$\chi^2_{.99}$	$\chi^2_{.995}$
1	.000039	.00016	.00098	.0039	.0158	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.1026	.2107	4.61	5.99	7.38	9.21	10.60
3	.0717	.115	.216	.352	.584	6.25	7.81	9.35	11.34	12.84
4	.207	.297	.484	.711	1.064	7.78	9.49	11.14	13.28	14.86
5	.412	.554	.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	.676	.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
120	83.85	86.92	91.58	95.70	100.62	140.23	146.57	152.21	158.95	163.64

For large degrees of freedom,

$$\chi^2_P = \tfrac{1}{2}(z_P + \sqrt{2v - 1})^2 \text{ approximately,}$$

where v = degrees of freedom and z_P is given in Table 2.