Swarthmore College
Statistics Honors Exam
Wednesday May 15, 1996

Instructions: This exam has three parts. Part I consists of elementary problems and part II includes problems that are more theoretical in nature. In part III the problems are more open ended, try to give a concise and carefully thought out discussion of the issues.

Answer the questions carefully, showing necessary work. Careful answers to most of the questions are more important than careless answers to all of them. However, be sure to answer *some questions from each part,* so do not spend to much time working on any one part. Make sure that you leave plenty of time for the discussions in part III. You may not have time to answer all of the questions.

Note: appropriate tables are attached

Part I:

1. Suppose a coin is tossed 100 times and heads comes up 60 times. Should we be surprised and perhaps doubt that the coin is fair?

2. For a Poisson random variable ($P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, \ldots$) determine the mean and variance. (Show your work for the derivation, don't just state the result.)

2. let X and Y be independent random variables with variances σ_X^2 and σ_Y^2 respectively. Let $U = X+Y$ and $V = X-Y$. Determine the covariance $\text{Cov}(U, V)$.

3. Let $X_1, X_2, \ldots, X_n$ be independent identically distributed random normal random variables with mean μ and variance σ^2. What are the method of moments estimates of μ and σ^2?

4. Is there any difference between the type of data analyzed in Chi-square test of homogeneity and analysis of variance (ANOVA) tests? What sorts of hypotheses are being tested in each case?

5. What are the different assumptions made about underlying distributions in 'standard' methods and nonparametric methods (such as Mann-Whitney, signed rank test etc).

6. Show that the exponential density ($f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$) has the memoryless property $P(X > t + s | X > s) = P(X > t)$.

Part II.

1. If X is a nonnegative continuous random variable, show that

$$E[X] = \int_0^\infty [1 - F(x)]dx$$

where $F(x)$ is the cumulative distribution function.

2. Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution having unknown mean μ and known variance σ^2.

(a) Determine a $100(1-\alpha)\%$ confidence interval for μ. Your answer will be in terms of α, $z(*)$, σ, $\overline{X} = (X_1 + \cdots + X_n)/n$ and/or n where $z(*)$ satisfies $P(Z > z(*)) = *$ for standard normal Z.

(b) Consider testing the following hypotheses:

$$H_0 : \mu = \mu_0 \qquad\qquad H_A : \mu \neq \mu_0.$$

For a given α determine x_0 so that H_0 is rejected (at level α) for $|\overline{X} - \mu_0| > x_0$.

(c) Discuss the relation between your answers to (a) and (b).

3. Let $X_1, X_2, \ldots, X_n$ be independent identically distributed Poisson random variables (i.i.d.) with parameter λ. (So $P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, \ldots$.) Determine the the maximum likelihood estimate of λ.

4. Let $X_1, X_2, \ldots, X_n$ be independent random variables with common distribution F and density f. Let $U = \max(X_1, X_2, \ldots, X_n)$ denote the maximum of the X_i. Show that the density of U is $f_U(u) = nf(u)[F(u)]^{n-1}$.

5. If $X_1, \ldots, X_n$ are i.i.d. uniform on $[0, \theta]$, the maximum likelihood estimate of θ is $\max(X_1, X_2, \ldots, X_n)$. Show that this is a biased estimate.

6. Consider fitting the straight line $y = \beta x$ to points $(x_i, y_i$ for $i = 1, 2, \ldots, n$. (Note there is no constant here as in the standard least square fit.) Write down the expression we would seek to minimize for the method of least squares. Determine the estimate of β in this model.

Part III

1. What is a random variable? Give as precise a definition as you can. Give also an informal description; one that you might give to a friend who has little background in mathematics.

2. What are bootstrap methods? Discuss, giving examples of where they might be used.

3. State the Central Limit Theorem. Give both an informal statement (one that you might give to a friend with no mathematical background) and as precise a mathematical statement as you can.

What is the importance of the Central Limit Theorem in determining a confidence interval for a mean? Do you need to make any assumptions about the underlying population distribution for your sample?

4. Recall that a Type I error involves rejecting the null hypothesis H_0 when it is true and a Type II error involves accepting the null hypothesis H_0 when it is false.

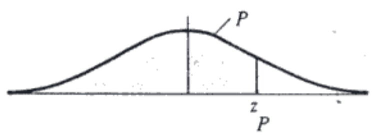
Which of these relates to the *significance level* and which to the *power* and how? Which level is usually decided on by the statistician?

Discuss what the null and alternative hypotheses are and why there is no simple relation between the probability of a Type I error and the probability of a Type II error.

Discuss anything else that you think is important in understanding the distinction between these two types of error, for example what the 'real world' implications might be in a typical hypothesis testing situation.

5. The Chi-square distribution (χ^2), the t distribution and the F distribution are all related to normal distributions. Describe what you remember about these distributions and their relationship to normal distributions. Also describe some of their uses in hypothesis testing.

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