Endterm

SDS321

Spring 2016 PART I

You may use two pages (4 sides) of notes, and you may use a calculator. You must hand in your notes, plus any rough work, after the exam.

This exam is one hour and fifteen minutes long. It contains 3 short (total 7 points) and 3 long (6 points each) questions.

Please answer all problems in the space provided on the exam. Use extra pages if needed. Of course, please put your name on extra pages.

Read each question carefully, show your work and clearly present your answers. Note, the exam is printed two-sided - please don't forget the problems on the even pages!

Good Luck!

Name: ____________________

UTeid: ____________________

1. (2 pts) The wait time X to get to the front of the queue in the doctor's office is distributed as $X \sim Exponential(Y)$. Y itself is uniformly distributed between one and two. What is the expected wait time, i.e. $E[X]$?

Solution: $E[X] = E[E[X|Y]] = E[1/Y] = \int_1^2 1/y dy = \log 2$

2. Let X be the number of heads from tosses of 10 fair coins.

   (a) (2 pts) Calculate $P(X \geq 9)$ exactly.

   Solution: $P(X \geq 9) = P(X = 9) + P(X = 10) = (10 + 1).5^10 = 11/2^{10} = .01$

   (b) (3 pts) Calculate an upper bound on this using the Chebyshev inequality. You can use the fact that the Binomial(n,1/2) PMF is symmetric around its expectation.

   Solution: $P(X \geq 9) = .5(P(|X - 5| \geq 4) \leq .5 \times 10 \times .25/16 = 1.25/16 = .08$

3. (6 pts) A plant grows in two distant islands. Suppose that the lifespan measured in days on the first island is normally distributed with unknown mean μ_X and known variance $\sigma_X^2 = 16$. Similarly the lifespan in the second island is normally distributed with unknown mean μ_Y and known variance $\sigma_Y^2 = 36$. We want to test the null hypothesis $H_0 : \mu_X = .5\mu_Y$. The alternative is $H_1 : \mu_X \neq .5\mu_Y$ based on a hundred independent samples from each island. To be specific, you see $x_1, \ldots, x_{100}$ from island one and $y_1, \ldots, y_{100}$ from island two. All these samples are independent of each other. The sample means are $\frac{\sum_i x_i}{100} = 48$ and $\frac{\sum_j y_j}{100} = 100$.

(a) (1 pt) Construct a sample statistic which has zero mean under the null hypothesis.

Solution: $S = \bar{X} - .5\bar{Y}$.

(b) (2 pts) What is its variance under the null?

Solution: $var(S) = 16/100 + 9/100 = .25$.

(c) (2 pts) Construct a rejection region at 95% significance level.

Solution: $P(|S| \geq \xi; H_0) = .05$. *Under the null,* $S \sim N(0, .25)$. *So* $\xi/.5 = 1.96$ *and so* $\xi = .98$.

(d) (1 pt) Would you accept or reject the null hypothesis at the 95% significance level?

Solution: $|S| = |48 - 50| = 3$. *Clearly in the rejection region. reject*

4. (6 pts) You have tossed 10 fair coins. Let X be the number of heads and Y be the number of tails.

   (a) (1 pt) Write Y in terms of X.

   Solution: $Y = 10 - X$.

   (b) (2 pts) What is the covariance of $2X$ and Y? *Hint: might be useful to use the properties of covariance we learned in class.*

   Solution: $cov(2X, (1 - X)) = 2cov(X, 1 - X) = -2var(X)$

   (c) (2 pts) What is the correlation coefficient of $2X$ and Y?

   Solution: $\rho(2X, (1-X)) = cov(2X, 1-X)/\sqrt{var(2X)var(1 - X)} = -1$

   (d) (1 pt) Explain why this makes sense in two sentences. Because $Y = (n - X)$, so for $V = 2X$, $Y = n - V/2$. Which means they are linearly related, and when V increases Y decreases. Which is why the corr. coeff. is -1.

5. (6 pts) Consider n independent random variables $X_1, X_2, \ldots, X_n \sim Geometric(p)$.

(a) (2 pts) Write down the likelihood.

$$\prod_i P(X_i; p) = p(1-p)^{X_i - 1}$$

.

(b) (3 pts) What is the maximum likelihood estimate of p?

Solution: *Log likelihood*

$$\ell(p) = \sum_i \log P(X_i; p) = n \log p + \sum_i (X_i - 1) \log(1-p)$$

. *Take derivative w.r.t* p. $n/p - (\sum_i (X_i - 1)/(1-p)$ *Set this to zero and solve for* $\hat{p}$.

$$\frac{\hat{p}}{1-\hat{p}} = \frac{n}{\sum_i X_i - n}$$

Solving

$$\hat{p} = \frac{n}{\sum_i X_i}$$

(c) (1 pt) Explain in a sentence why this makes sense.

Solution: Because we know that $E[X] = 1/p$. *So* $E[1/\hat{p}] = E[X_1] = 1/p$. *This is again very similar to the exponential distribution.*

Useful distributions

All PDFs/PMFs are zero outside the range specified.

	PDF/PMF	$E[X]$	$\mathrm{var}(X)$
Bernoulli	$p^x(1-p)^{1-x}, x = 0, 1$	p	$p(1-p)$
Binomial	$\binom{n}{k}p^k(1-p)^{n-k}$, k=0,1,...,n	np	$np(1-p)$
Geometric	$p(1-p)^{k-1}, k = 1, 2, 3, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Poisson	$\frac{\lambda^k e^{-\lambda}}{k!}, k = 1, 2, 3, \ldots$	λ	λ
Normal	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Exponential	$\lambda e^{-\lambda x}, x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Uniform	$\frac{1}{b-a} \qquad a \leq x \leq b, b > a$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$

Standard normal table

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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	0.8413	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	0.9772	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	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990