# Midterm Exam 2

Math 20

August 9, 2012

Name: Answer Key

**Instructions:** This exam is closed-book, with no calculators, notes, or books allowed. You may not give or receive any help on the exam, though you may ask the instructor for clarification if necessary. Be sure to show all your work wherever possible. You can leave your answer in terms of factorials, binomial coefficients, fractions, exponentials, etc. unless explicitly stated otherwise. The normal distribution table is located on the last page.

HONOR STATEMENT: I have neither given nor received any help on this exam, and I attest that all of the answers are my own work.

SIGNATURE: ____________________

| Problem | Score | Points |
|---|---|---|
| 1 | | 12 |
| 2 | | 14 |
| 3 | | 18 |
| 4 | | 5 |
| 5 | | 12 |
| 6 | | 11 |
| 7 | | 12 |
| 8 | | 16 |
| Total | | 100 |

1. [2 points each] **Fill in the Blank.** In the blank space, write the name of the distribution described.

(a.) A random variable $X$ has the geometric distribution if $X$ is the number of Bernoulli trials until you get the first success. If $p$ is the probability of success, then the expected value of $X$ is $\frac{1}{p}$ and the variance is $\frac{1-p}{p^2}$.

(b.) A random variable $X$ has the uniform distribution if every value of $X$ is equally likely. If the values $X$ takes are the positive integers up to $n$, then the expected value of $X$ is $\frac{n+1}{2}$ and the variance of $X$ is $\frac{n^2-1}{2}$.

(c.) A random variable $X$ has the Poisson distribution (approximately) if $X$ counts the number of occurrences of some event (or successes) within a given interval of time or space. Usually you are given an average rate $\lambda$. The expected value of $X$ is $\lambda$ and the variance of $X$ is $\lambda$.

(d.) A random variable $X$ has the Negative Binomial distribution if $X$ is the number of Bernoulli trials until you get the $k$th success. The expected value of $X$ is $\frac{k}{p}$ and the variance of $X$ is $k\left(\frac{1-p}{p^2}\right)$.

(e.) A random variable $X$ has the Binomial distribution if $X$ is the number of successes in a Bernoulli trial process. If you perform $n$ trials, each with probability $p$ of success, then the expected value of $X$ is $np$ and the variance of $X$ is $npq$.

(f.) A random variable $X$ has the hypergeometric distribution if $X$ counts the number of successes in $n$ trials done without replacement (so the trials are not independent). If $N$ is the population and there are $m$ successes in the population, then the expected value of $X$ is $\frac{nk}{N}$ and the variance of $X$ is $np(1-p)\left(1-\frac{n-1}{N-1}\right)$.

**2.** [2 points each] An urn contains 20 balls. Eight balls are red, seven balls are blue and the remaining 5 balls are yellow. The balls also have numbers written on them. The red balls are numbered 1-8, the blue balls are numbered 9-15, and the yellow balls are numbered 16-20. You don't need to provide any work for the following questions, only answers.

(a.) Draw 40 balls with replacement. What is the probability you draw exactly 20 red balls?

$$b\left(40, \frac{8}{20}, 20\right) = \binom{40}{20}\left(\frac{8}{20}\right)^{20}\left(\frac{12}{20}\right)^{20}$$

(b.) What is the probability that you have to draw (with replacement) exactly five times to get your first yellow ball?

$$\left(\frac{15}{20}\right)^{4}\left(\frac{5}{20}\right)$$

(c.) What is the expected number of draws (with replacement) you have to make to get 3 blue balls with even numbers on them?

20

(d.) What is the probability that it takes exactly 30 draws (with replacement) in order to get 3 blue balls with even numbers on them? (That is, the 30th draw is the third blue ball with an even number on it.)

$$\binom{29}{2}\left(\frac{3}{20}\right)^3\left(\frac{17}{20}\right)^{27}$$

(e.) Suppose you draw four balls *with* replacement. What is the probability that you draw at least one red ball?

$$1-\binom{4}{0}\left(\frac{8}{20}\right)^0\left(\frac{12}{20}\right)^4$$

(f.) What is the expected number of red balls you get when you draw 5 balls with replacement?

2

(g.) What is the expected number of red balls you get when you draw 5 balls without replacement?

2

**3.** [6 points each] Answer the following questions:

(a.) Suppose you have independent identically distributed random variables $X_1, X_2, \cdots, X_n$ with mean $\mu$ and variance $\sigma^2$. How big must $n$ be to guarantee that there is at least a 99% chance that the sample mean will be within two standard deviations of the true mean $\mu$?

Use Chebyshev's Inequality.

$$P(|X-\mu| \geq a) \leq \frac{V(x)}{a^2}$$

$$P\left(\left|\frac{S_n}{n}-\mu\right| \geq 2\sigma\right) \leq \frac{V\left(\frac{S_n}{n}\right)}{(2\sigma)^2} = \frac{\sigma^2}{n} \cdot \frac{1}{4\sigma^2} = \frac{1}{4n}$$

Want $P\left(\left|\frac{S_n}{n}-\mu\right| \geq 2\sigma\right) \leq .01$, so we need $\boxed{n=25}$

(just solve for $n$)

(b.) A restaurant has 87 tables, so it can seat 87 groups for dinner. The manager knows from experience that groups only have an 80% chance of showing up when they have made a reservation. Therefore, the manager often accepts more reservations than he has tables. Suppose that he takes 100 reservations. Estimate the probability that the restaurant cannot seat everyone that shows up. (That is, estimate the probability that more than 87 groups with reservations show up.)

Let $X_i = \begin{cases} 1 & \text{if } i^{th} \text{ group shows up} \\ 0 & \text{otherwise.} \end{cases}$ $S_n =$ # groups that show up.

Want $P(S_n > 87) = P(S_n \geq 88) = P\left(S_n^* \geq \frac{88-100(.8)}{\sqrt{100(.8)(.2)}}\right)$

$$= P\left(S_n^* \geq \frac{88-80}{\sqrt{16}}\right) = P\left(S_n^* \geq \frac{8}{4}\right) = P(S_n^* \geq 2)$$

$$= \int_2^\infty \varphi(x)dx = .5 - .4772 = \boxed{.0228}$$

(c.) Suppose you are conducting a poll. How big should $n$ be to get a 1% margin of error (or 1% accuracy) for the 95% confidence interval? (Note that 1% margin of error means that your confidence interval will have length 0.02.)

Want 1% accuracy, so $2\sigma \le .01$

standard deviation of $\bar{p}$.

So want $2\sqrt{\frac{pq}{n}} \le .01$

Recall: $pq \le \frac{1}{4}$ always, so $2\sqrt{\frac{1/4}{n}} = \frac{1}{\sqrt{n}} \le .01$

$\Rightarrow \frac{1}{\sqrt{n}} \le \frac{1}{100}$ so need $n \ge 10{,}000$.

4. [5 points] State the Law of Large Numbers.

Let $X_i$'s be identically distributed independent random variables w/ $E(X_i) = \mu$.

Let $S_n = X_1 + X_2 + \cdots + X_n$.

Then if $a > 0$,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \le a\right) \to 1 \quad \text{as } n \to \infty$$

and $$P\left(\left|\frac{S_n}{n} - \mu\right| \ge a\right) \to 0 \quad \text{as } n \to \infty$$

5. **Proofs.** Show all steps.

(a.) [6 points] Use Markov's inequality to prove Chebyshev's inequality. (You should state Markov's inequality and note where you use it in your proof.)

Markov's Inequality: $P(|X| \geq a) \leq \frac{E(|X|)}{a}$ for positive $a$.

$$P(|X-\mu| \geq a) = P((X-\mu)^2 \geq a^2) \underset{\text{Markov's}}{\leq} \frac{E((X-\mu)^2)}{a^2} = \frac{V(X)}{a^2}.$$

This gives us Chebyshev's Inequality:

$P(|X-\mu| \geq a) \leq \frac{V(X)}{a^2}$ for positive $a$.

(b.) [6 points] Prove that $E(\frac{1}{X}) \geq \frac{1}{E(X)}$, where $X$ is a positive random variable.

Use Jensen's inequality.

$f(x) = \frac{1}{x}$

$f'(x) = -\frac{1}{x^2}$

$f''(x) = \frac{2}{x^3} > 0$ when $x$ is positive. So $f$ is convex.

Therefore $E(f(x)) \geq f(E(X)) \Rightarrow E(\frac{1}{X}) \geq \frac{1}{E(X)}$.

6. Suppose there are a million people living in a city. During each day, each person has probability $p = 2 \times 10^{-6}$ of calling the fire station.

(a.) [7 points] Use Poisson approximation to estimate the probability that on a particular day, the fire station gets at least two calls.

$$\lambda = (2 \cdot 10^{-6}) \cdot (10^{6}) = 2$$

$$P(X \geq 2) = 1 - P(X=0) - P(X=1)$$

$$= 1 - \frac{2^0}{0!} e^{-2} - \frac{2^1}{1!} e^{-2} = \boxed{1 - 3e^{-2}}$$

(b.) [4 points] Approximately how many days will it take on average until the fire station gets at least two calls in the same day (including the day this happened)?

Geometric dist, so $\frac{1}{p} = \boxed{\frac{1}{1-3e^{-2}}}$

7. A surveying instrument makes an error of -2, -1, 0, 1, or 2 feet with equal probabilities when measuring the height of a 200-foot tower.

(a.) [4 points] Find the expected value and variance for the height obtained by using this instrument once.

$X$ = height obtained
$Y$ = error. Then $X = 200 + Y$.

$E(Y) = \frac{1}{5}(-2 + -1 + 0 + 1 + 2) = 0$, So $E(X) = E(200+Y)$
$= 200 + E(Y) = \boxed{200}$

$V(Y) = \frac{1}{5}((-2)^2 + (-1)^2 + 0^2 + 1^2 + 2^2) = \frac{10}{5} = 2.$

So $V(X) = V(200+Y) = V(Y) = \boxed{2}$

(b.) [8 points] Estimate the probability that in 18 independent measurements of this tower, the average of the measurements is between 199 and 201, inclusive.

$S_{18} = X_1 + X_2 + \cdots + X_{18}$. Average height $= \frac{S_n}{n}$

$E\left(\frac{S_n}{n}\right) = 200$
$V\left(\frac{S_n}{n}\right) = \frac{1}{n} V(X_i) = \frac{1}{18} \cdot 2 = \frac{1}{9}$

$P\left(199 \le \frac{S_n}{n} \le 201\right) = P\left(\frac{199-200}{\sqrt{1/9}} \le \left(\frac{S_n}{n}\right)^* \le \frac{201-200}{\sqrt{1/9}}\right)$

$= P\left(-3 \le \left(\frac{S_n}{n}\right)^* \le 3\right) = 2(.4987) = \boxed{.9974}$

8. [16 points] For this problem, you should give the complete statement of two different versions of the Central Limit Theorem (there were four versions that we talked about in class). You should define/explain any notation you use (random variables, functions, etc.). In addition, for each version, you should give an example of the type of question that version is meant to answer and justify why this example would best be solved using the corresponding version of the Central Limit Theorem. (If you need more space, the next page is blank for your use.)

Version 1: For the binomial distribution $b(n,p,j)$,

$$\lim_{n\to\infty} \sqrt{npq}\, b(n,p,j) = \varphi(j^*)$$

where $j^* = \frac{j-np}{\sqrt{npq}}$ and $\varphi(x)$ is the standard normal density function, defined $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

As usual, $p$ = prob of success

$q$ = prob of failure.

$j$ = # of successes

$n$ = # of trials.

Ex. Estimate probability of getting exactly 52 heads when flipping a coin 100 times, b/c here we want to estimate $b(100, \frac{1}{2}, 52)$.

---

Version 2: Let $S_n$ = # of successes in $n$ Bernoulli trials w/ prob $p$ for success. Let $a$ & $b$ be two fixed #s w/ $a<b$. Then

$$\lim_{n\to\infty} P(a \le S_n \le b) = \int_{a^*}^{b^*} \varphi(x)\,dx$$

where $a^* = \frac{a-np}{\sqrt{npq}}$, $b^* = \frac{b-np}{\sqrt{npq}}$.

defined as above.

$\Rightarrow$

Example: Probability that when you flip a coin 100 times, you get btwn 40 & 60 heads, because you're dealing w/ a range of values.

---

Version 3: Let $S_n = X_1 + \cdots + X_n$ be the sum of n discrete ind. identically distributed r.v.s w/ $E(X_i) = \mu$, $V(X_i) = \sigma^2$. Then for $a<b$,

$$\lim_{n\to\infty} P(a \le S_n \le b) = \int_{a^*}^{b^*} \varphi(x)\,dx$$

as above.

where $a^* = \dfrac{a - n\mu}{\sqrt{n\sigma^2}}$, $b^* = \dfrac{b-n\mu}{\sqrt{n\sigma^2}}$.

Ex: Prob that the sum of 100 dice is btwn 350 & 360, b/c this is not a Bernoulli trial, but is sum of independent identically distributed r.v.s.

---

Version 4: $X_1, X_2, \ldots, X_n$ independent. $S_n = X_1 + \cdots + X_n$. $E(X_i) = \mu_i$, $V(X_i) = \sigma_i^2$. Then $E(S_n) = m_i = \mu_1 + \cdots + \mu_n$. $V(S_n) = S_n^i = \sigma_1^2 + \cdots + \sigma_n^2$. Assume $S_n^2 \to \infty$ as $n \to \infty$. If there is some constant $A$ s.t. $|X_n| \le A$ for all n, then

$$\lim_{n\to\infty} P(a \le S_n \le b) = \int_{a^*}^{b^*} \varphi(x)\,dx \quad \text{where } a^* = \frac{a - m_n}{S_n}, \quad b^* = \frac{b - m_n}{S_n}.$$



Ex: Use to show why height is normally distributed b/c involves many genes/factors, not identical.

**Bonus:** Let $X$ have the Poisson Distribution with parameter $\lambda > 0$. Find $E(X!)$, the expected value of $X$ factorial. You answer should be simplified with no summation sign.

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

$$E(X!) = \sum_{k=0}^{\infty} k!\, P(X=k) = \sum_{k=0}^{\infty} k! \left(\frac{\lambda^k}{k!} e^{-\lambda}\right)$$

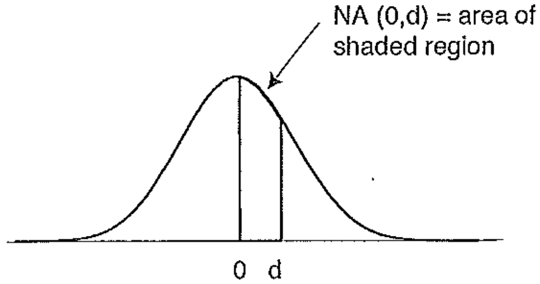
$$= \sum_{k=0}^{\infty} \lambda^k e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \lambda^k.$$

So if $\lambda \geq 1$, $E(X!) = \infty$

if $\lambda < 1$, $E(X!) = \dfrac{e^{-\lambda}}{1-\lambda}$

# Appendix A

## Normal distribution table




| | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .0000 | .0040 | .0080 | .0120 | .0160 | .0199 | .0239 | .0279 | .0319 | .0359 |
| 0.1 | .0398 | .0438 | .0478 | .0517 | .0557 | .0596 | .0636 | .0675 | .0714 | .0753 |
| 0.2 | .0793 | .0832 | .0871 | .0910 | .0948 | .0987 | .1026 | .1064 | .1103 | .1141 |
| 0.3 | .1179 | .1217 | .1255 | .1293 | .1331 | .1368 | .1406 | .1443 | .1480 | .1517 |
| 0.4 | .1554 | .1591 | .1628 | .1664 | .1700 | .1736 | .1772 | .1808 | .1844 | .1879 |
| 0.5 | .1915 | .1950 | .1985 | .2019 | .2054 | .2088 | .2123 | .2157 | .2190 | .2224 |
| 0.6 | .2257 | .2291 | .2324 | .2357 | .2389 | .2422 | .2454 | .2486 | .2517 | .2549 |
| 0.7 | .2580 | .2611 | .2642 | .2673 | .2704 | .2734 | .2764 | .2794 | .2823 | .2852 |
| 0.8 | .2881 | .2910 | .2939 | .2967 | .2995 | .3023 | .3051 | .3078 | .3106 | .3133 |
| 0.9 | .3159 | .3186 | .3212 | .3238 | .3264 | .3289 | .3315 | .3340 | .3365 | .3389 |
| 1.0 | .3413 | .3438 | .3461 | .3485 | .3508 | .3531 | .3554 | .3577 | .3599 | .3621 |
| 1.1 | .3643 | .3665 | .3686 | .3708 | .3729 | .3749 | .3770 | .3790 | .3810 | .3830 |
| 1.2 | .3849 | .3869 | .3888 | .3907 | .3925 | .3944 | .3962 | .3980 | .3997 | .4015 |
| 1.3 | .4032 | .4049 | .4066 | .4082 | .4099 | .4115 | .4131 | .4147 | .4162 | .4177 |
| 1.4 | .4192 | .4207 | .4222 | .4236 | .4251 | .4265 | .4279 | .4292 | .4306 | .4319 |
| 1.5 | .4332 | .4345 | .4357 | .4370 | .4382 | .4394 | .4406 | .4418 | .4429 | .4441 |
| 1.6 | .4452 | .4463 | .4474 | .4484 | .4495 | .4505 | .4515 | .4525 | .4535 | .4545 |
| 1.7 | .4554 | .4564 | .4573 | .4582 | .4591 | .4599 | .4608 | .4616 | .4625 | .4633 |
| 1.8 | .4641 | .4649 | .4656 | .4664 | .4671 | .4678 | .4686 | .4693 | .4699 | .4706 |
| 1.9 | .4713 | .4719 | .4726 | .4732 | .4738 | .4744 | .4750 | .4756 | .4761 | .4767 |
| 2.0 | .4772 | .4778 | .4783 | .4788 | .4793 | .4798 | .4803 | .4808 | .4812 | .4817 |
| 2.1 | .4821 | .4826 | .4830 | .4834 | .4838 | .4842 | .4846 | .4850 | .4854 | .4857 |
| 2.2 | .4861 | .4864 | .4868 | .4871 | .4875 | .4878 | .4881 | .4884 | .4887 | .4890 |
| 2.3 | .4893 | .4896 | .4898 | .4901 | .4904 | .4906 | .4909 | .4911 | .4913 | .4916 |
| 2.4 | .4918 | .4920 | .4922 | .4925 | .4927 | .4929 | .4931 | .4932 | .4934 | .4936 |
| 2.5 | .4938 | .4940 | .4941 | .4943 | .4945 | .4946 | .4948 | .4949 | .4951 | .4952 |
| 2.6 | .4953 | .4955 | .4956 | .4957 | .4959 | .4960 | .4961 | .4962 | .4963 | .4964 |
| 2.7 | .4965 | .4966 | .4967 | .4968 | .4969 | .4970 | .4971 | .4972 | .4973 | .4974 |
| 2.8 | .4974 | .4975 | .4976 | .4977 | .4977 | .4978 | .4979 | .4979 | .4980 | .4981 |
| 2.9 | .4981 | .4982 | .4982 | .4983 | .4984 | .4984 | .4985 | .4985 | .4986 | .4986 |
| 3.0 | .4987 | .4987 | .4987 | .4988 | .4988 | .4989 | .4989 | .4989 | .4990 | .4990 |
| 3.1 | .4990 | .4991 | .4991 | .4991 | .4992 | .4992 | .4992 | .4992 | .4993 | .4993 |
| 3.2 | .4993 | .4993 | .4994 | .4994 | .4994 | .4994 | .4994 | .4995 | .4995 | .4995 |
| 3.3 | .4995 | .4995 | .4995 | .4996 | .4996 | .4996 | .4996 | .4996 | .4996 | .4997 |
| 3.4 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4998 |
| 3.5 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 |
| 3.6 | .4998 | .4998 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 |
| 3.7 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 |
| 3.8 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 | .4999 |
| 3.9 | .5000 | .5000 | .5000 | .5000 | .5000 | .5000 | .5000 | .5000 | .5000 | .5000 |