Preparing for the <u>Midterm Exam</u>

- Thursday, November 9, 2pm-3:20pm
- Here in 305 MSE
- Closed book exam; no books, notes, laptops, etc.
- However, calculators can be used.
- You can prepare one cheat sheet (letter size)

This will be provided

Name	Probability Distribution	Mean	Variance	Section in Book
Discrete				
Uniform	$\frac{1}{n}, a \le b$	$\frac{(b+a)}{2}$	$\frac{(b-a+1)^2-1}{12}$	3-5
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n, 0 \le p \le 1$	np	$np(1-p)$	3-6
Geometric	$(1-p)^{x-1} p$, $x = 1, 2, \ldots, 0 \le p \le 1$	$1/p$	$(1-p)/p^2$	3-7.1
Negative binomial	$\binom{x-1}{r-1} (1-p)^{x-r} p^r$, $x = r, r+1, r+2, \ldots, 0 \le p \le 1$	r/p	$r(1-p)/p^2$	3-7.2
~~Hypergeometric~~	$\cancel{\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}}$, $\cancel{x = \max(0, n-N+K), 1}$ $\cancel{\min(K, n), K \le N, n \le N}$	$\cancel{np}$, $\cancel{\text{where } p = \frac{K}{N}}$	$\cancel{np(1-p)\left(\frac{N-n}{N-1}\right)}$	~~3-8~~
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots, 0 < \lambda$	λ	λ	3-9
Continuous				
Uniform	$\frac{1}{b-a}, a \le x \le b$	$\frac{(b+a)}{2}$	$\frac{(b-a)^2}{12}$	4-5
Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, $-\infty < x < \infty, -\infty < \mu < \infty, 0 < \sigma$	μ	σ^2	4-6
Exponential	$\lambda e^{-\lambda x}, 0 \le x, 0 < \lambda$	$1/\lambda$	$1/\lambda^2$	4-8
Erlang	$\frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}, 0 < x, r = 1, 2, \ldots$	r/λ	r/λ^2	4-9.1
Gamma	$\frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)}, 0 < x, 0 < r, 0 < \lambda$	r/λ	r/λ^2	4-9.2

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

What may be on the exam?

- Bayes Theorem
- Discrete & Continuous Random Variables
- Joint Probability Distributions, Covariance/Correlations
- Confidence Intervals
- Hypothesis testing
- Other topics
- Look at Homeworks 1-4 for examples of problems

Hypothesis testing

3. (8 points) The college bookstore tells prospective students that the average cost of its textbooks is \$52 with a standard deviation of \$4.50. A group of statistics students think that the average cost is **actually higher**. In order to test bookstore's claim against this alternative hypothesis, the students bought a random sample of 100 books. The mean price of this sample was \$52.80. Perform the hypothesis test at the 5% level of significance and state your decision.

What type of hypothesis should I apply?

A. Two-sided: $\mu_1 \neq \mu_0$
B. $\boxed{\text{One-sided: } \mu_1 > \mu_0}$
C. One-sided: $\mu_1 < \mu_0$
D. Three-sided
E. I have no idea

Get your i-clickers

3. **(8 points)** The college bookstore tells prospective students that the average cost of its textbooks is \$52 with a standard deviation of \$4.50. A group of statistics students think that the average cost is **actually higher**. In order to test bookstore's claim against this alternative hypothesis, the students bought a random sample of 100 books. The mean price of this sample was \$52.80. Perform the hypothesis test at the 5% level of significance and state your decision.

The standard deviation of $\bar{x}$ in this sample is:

A. \$4.50
B. \$45
C. $\boxed{\text{\$0.45}}$
D. I have no idea

Get your i-clickers

3. **(8 points)** The college bookstore tells prospective students that the average cost of its textbooks is \$52 with a standard deviation of \$4.50. A group of statistics students think that the average cost is **actually higher**. In order to test bookstore's claim against this alternative hypothesis, the students bought a random sample of 100 books. The mean price of this sample was \$52.80. Perform the hypothesis test at the 5% level of significance and state your decision.

3. (8 points) The college bookstore tells prospective students that the average cost of its textbooks is \$52 with a standard deviation of \$4.50. A group of statistics students think that the average cost is **actually higher**. In order to test bookstore's claim against this alternative hypothesis, the students bought a random sample of 100 books. The mean price of this sample was \$52.80. Perform the hypothesis test at the 5% level of significance and state your decision.

Answer: Hypothesis: $\begin{Bmatrix} H_0 : \mu = 52 \\ H_1 : \mu > 52 \end{Bmatrix}$. The critical z-value can be obtained from $z^* = \dfrac{52.8 - 52}{4.5/10} = 1.78$ Since $z^* > z_\alpha = 1.65$, this test statistic lies in the rejection region for H_0. Thus null hypothesis H_0 will be rejected and alternative hypothesis H_1 is accepted.

Confidence intervals

1. A sample of size 100 which has the sample mean $\bar{X} = 500$ was drawn from a population with an unknown mean μ and the standard deviation $\sigma = 80$.

a) Give the 95% confidence interval for the population mean.

1. A sample of size 100 which has the sample mean $\bar{X} = 500$ was drawn from a population with an unknown mean μ and the standard deviation $\sigma = 80$.

a) Give the 95% confidence interval for the population mean.

Answer: $$P(500 - z_{0.025}\frac{80}{\sqrt{100}} < Z < 500 + z_{0.025}\frac{80}{\sqrt{100}}) = 0.95$$ where $z_{0.025} = 1.96$.

Therefore, the interval is [484.32, 515.68].

Bayes rule

In answering a question on a multiple-choice test, a student either knows the answer or he guesses. Let 1/3 be the probability that he knows the answer. If he does not know the answer, he randomly guesses one out of 4 multiple choice questions. What is the conditional probability that a student knew the answer to a question given that he answered it correctly?

A. 1/4
B. 1/3
C. $\boxed{\text{2/3}}$
D. 1/5
E. I don't know

Get your i-clickers

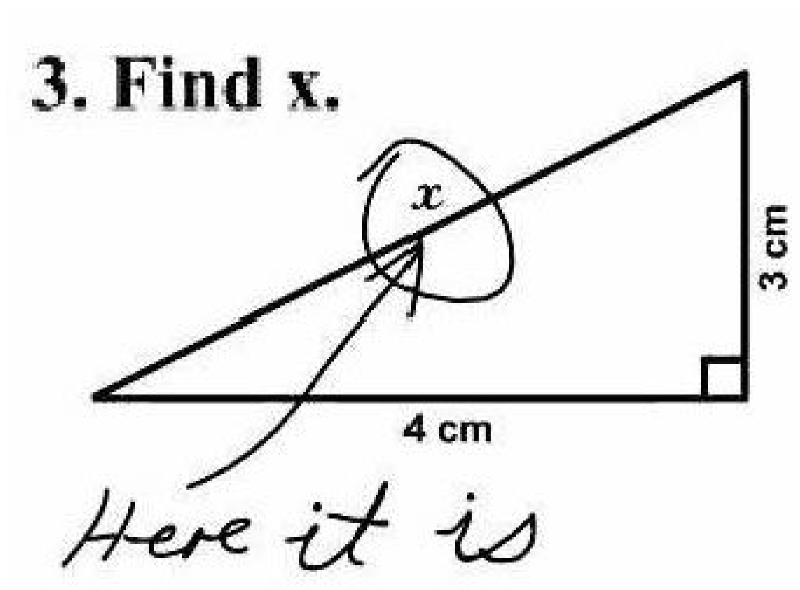
In answering a question on a multiple-choice test, a student either knows the answer or he guesses. Let 1/3 be the probability that he knows the answer. If he does not know the answer, he randomly guesses one out of 4 multiple choice questions. What is the conditional probability that a student knew the answer to a question given that he answered it correctly?

Answer: $P(K)=1/3$, $P(K')=2/3$, $P(C|K)=1$, $P(C|K')=1/4$.
$P(K|C)=P(C|K)*P(K)/P(C)=1*(1/3)/(1*1/3+(1/4)*(2/3))=2/3=0.666\ldots$

Discrete Probability Distributions

What is X in this problem?

- What is the random variable: Look for keywords:
  - Find the probability that....
  - What is the mean (or variance) of…
- What are parameters? Look for keywords:
  - Given that…
  - Assuming that…



Guide to probability distributions

- Binomial: # of samples, n, is fixed, # of successes, x, is variable

$$P(X=x)=\frac{n!}{x!(n-x)!}p^{x}(1-p)^{n-x}$$

- Geometric: # of samples, x is variable. # of successes 1 is fixed. Success comes in the end

$$P(X=x)=(1-p)^{x-1}\cdot p$$

- Negative binomial: # of samples, x is variable. # of successes, r, is fixed. rth success in ~~the~~ end

$$P(X=x)=\frac{(x-1)!}{(r-1)!(x-r)!}p^{r}(1-p)^{x-r}$$

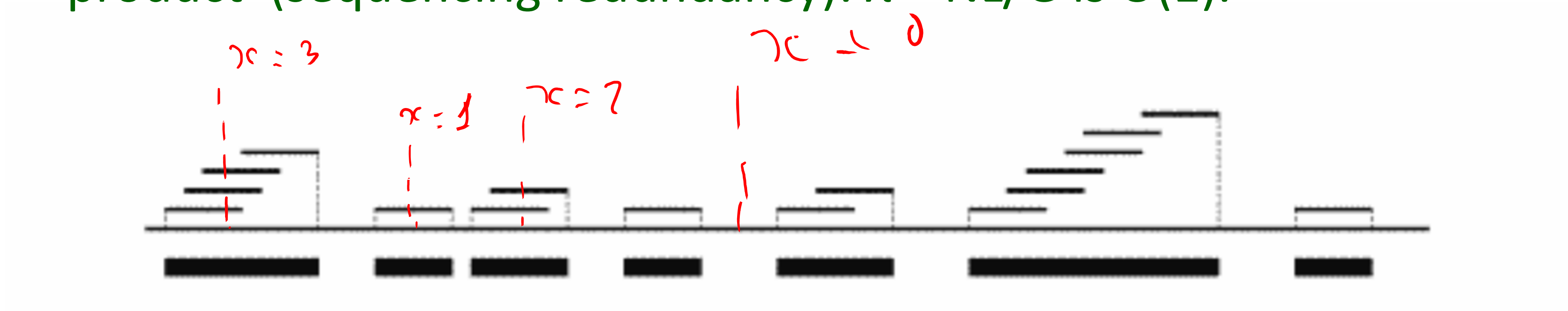
Poisson distribution in genomics

- *G - genome length (in bp)*
- *L - short read average length*
- *N – number of short read sequenced*
- *λ – sequencing redundancy = LN/G*
- *x- number of short reads covering a given site on the genome*

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

Ewens, Grant, Chapter 5.1

Poisson as a limit of Binomial. For a given site on the genome for each short read Prob(site covered): p=L/G is very small. Number of attempts (short reads): N is very large. Their product (sequencing redundancy): λ = NL/G is O(1).



Probability that a base pair in the genome is not covered by any short reads is 0.1

One randomly selects base pairs until exactly 5 uncovered base pairs are found.

Which discrete probability distribution describes the number of attempts?

A. Poisson
B. Binomial
C. Geometric
D. Negative Binomial
E. I have no idea

Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots, 0 < \lambda$
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, \ldots, n, 0 \le p \le 1$
Geometric	$(1-p)^{x-1} p$ $x = 1, 2, \ldots, 0 \le p \le 1$
Negative binomial	$\binom{x-1}{r-1}(1-p)^{x-r} p^r$ $x = r, r+1, r+2, \ldots, 0 \le p \le 1$

Get your i-clickers

Probability that a base pair in the genome is not covered by any short reads is 0.1

One randomly selects base pairs until exactly 5 uncovered base pairs are found.

What are the values of p, r?

A. p=0.5, r=5
B. p=0.1, r=0.5
C. p=0.1, r=5
D. p=0.5, r=0.1
E. I have no idea

Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots, 0 < \lambda$
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$
	$x = 0, 1, \ldots, n, 0 \le p \le 1$
Geometric	$(1-p)^{x-1} p$
	$x = 1, 2, \ldots, 0 \le p \le 1$
Negative binomial	$\binom{x-1}{r-1} (1-p)^{x-r} p^r$
	$x = r, r+1, r+2, \ldots, 0 \le p \le 1$

Get your i-clickers

Cancer happens when the gene p53 mutates.
Probability of p53 to mutate per year is 5%.

How many years before a patient gets disease?

Which discrete probability distribution would you use to answer?

A. Poisson
B. Binomial
C. Geometric
D. Negative Binomial
E. I have no idea

Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots, 0 < \lambda$
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$
	$x = 0, 1, \ldots, n, 0 \le p \le 1$
Geometric	$(1-p)^{x-1} p$
	$x = 1, 2, \ldots, 0 \le p \le 1$
Negative binomial	$\binom{x-1}{r-1} (1-p)^{x-r} p^r$
	$x = r, r+1, r+2, \ldots, 0 \le p \le 1$

Get your i-clickers

Continuous Probability Distributions

2. (8 points) The length of stay at a specific emergency department in Phoenix, Arizona, in 2009 had a mean of 4.6 hours with a standard deviation of 2.9. Assume that the length of stay is normally distributed.

(A) (4 points) What is the probability of a length of stay greater than 10 hours?

(B) (4 points) How long does one have to stay in this emergency room to know that approximately 25% of all visits last even longer?

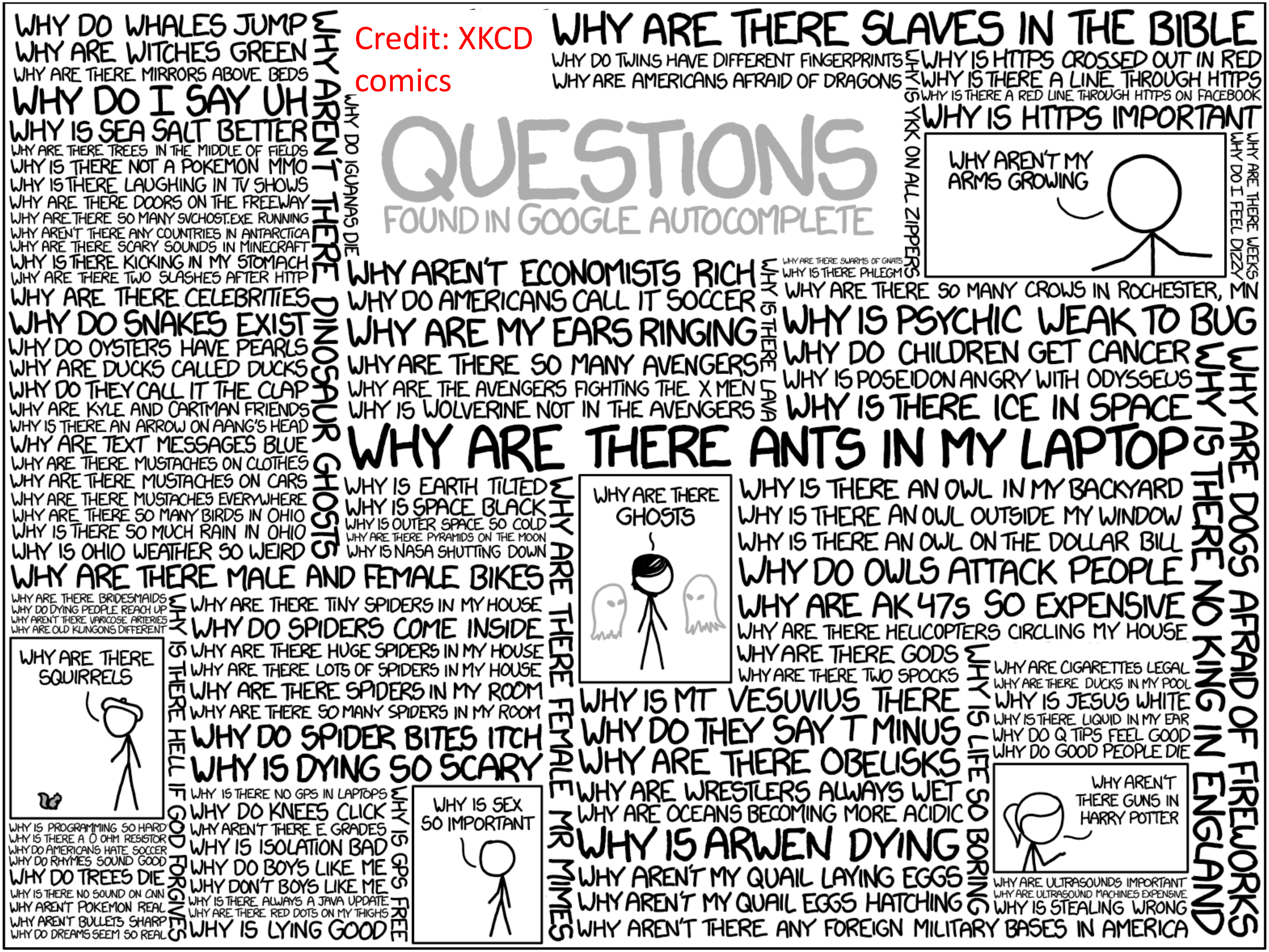
2. (8 points) The length of stay at a specific emergency department in Phoenix, Arizona, in 2009 had a mean of 4.6 hours with a standard deviation of 2.9. Assume that the length of stay is normally distributed.

(A) (4 points) What is the probability of a length of stay greater than 10 hours?

Answer: (10-4.6)/2.9=1.86 Using table one finds Prob=1-0.9687=0.0313

(B) (4 points) How long does one have to stay in this emergency room to know that approximately 25% of all visits last even longer?

Answer: Using table one finds P(Z<0.67)=0.75 meaning it is 4.6+2.9*0.67=6.543



Credit: XKCD comics

Two-sample hypothesis

Mating Calls. In a study of mating calls in the gray treefrogs *Hyla hrysoscelis* and *Hyla versicolor*, Gerhart (1994) reports that in a location in Lousiana the following data on the length of male advertisement calls have been collected:

	Sample size	Average duration	SD of duration	Duration range
Hyla chrysoscelis	43	0.65	0.18	0.36–1.27
Hyla versicolor	12	0.54	0.14	0.36–0.75

The two species cannot be distinguished by external morphology, but *H. chrysoscelis* are diploids while *H. versicolor* are tetraploids. The triploid

crosses exhibit high mortality in larval stages, and if they attain sexual maturity, they are sterile. Females responding to the mating calls try to avoid mismatches.

Based on the data summaries provided, test whether the length of call is a discriminatory characteristic? Use $\alpha = 0.05$.

	Sample size	Average duration	SD of duration
Hyla chrysoscelis	43	0.65	0.18
Hyla versicolor	12	0.54	0.14

Based on the data summaries provided, test whether the length of call is a discriminatory characteristic? Use $\alpha = 0.05$.

	Sample size	Average duration	SD of duration
Hyla chrysoscelis	43	0.65	0.18
Hyla versicolor	12	0.54	0.14

Based on the data summaries provided, test whether the length of call is a discriminatory characteristic? Use $\alpha = 0.05$.

1. Use two-sided hypothesis
2. z_{alpha/2}=1.96
3. Z=(0.65-0.54)/sqrt(0.18.^2/43+0.14.^2/12)=2.2516
4. Since Z> z_{alpha/2} null hypothesis can be rejected

Continuous probability distributions

Equipment Aging. Suppose that the lifetime T of a particular piece of laboratory equipment (in 1000 hour units) is an exponentially distributed random variable such that $\mathbb{P}(T > 10) = 0.8$.

(a) Find the "rate" parameter, λ.

(b) What are the mean and standard deviation of the random variable T?

Equipment Aging. Suppose that the lifetime T of a particular piece of laboratory equipment (in 1000 hour units) is an exponentially distributed random variable such that $\mathbb{P}(T > 10) = 0.8$.

(a) Find the "rate" parameter, λ.

(b) What are the mean and standard deviation of the random variable T?

PDF: $P(t) = \lambda \exp(-\lambda t)$

Inverse CDF: $P(T > t) = \exp(-\lambda t)$

$$-\lambda \cdot 10 = \ln(0.8)$$

$$\lambda = 0.0223$$