

Class 27: review for final exam –solutions, 18.05, Spring 2014

**This does not cover everything on the final.
Look at the posted practice problems for other topics.**

To save time in class: set up, but do not carry out computations.

Problem 1. (Counting)

(a) There are several ways to think about this. Here is one.

The 11 letters are p, r, o, b,b, a, i,i, l, t, y. We use the following steps to create a sequence of these letters.

Step 1: Choose a position for the letter p: 11 ways to do this.

Step 2: Choose a position for the letter r: 10 ways to do this.

Step 3: Choose a position for the letter o: 9 ways to do this.

Step 4: Choose two positions for the two b's: 8 choose 2 ways to do this.

Step 5: Choose a position for the letter a: 6 ways to do this.

Step 6: Choose two positions for the two i's: 5 choose 2 ways to do this.

Step 7: Choose a position for the letter l: 3 ways to do this.

Step 8: Choose a position for the letter t: 2 ways to do this.

Step 9: Choose a position for the letter y: 1 ways to do this.

Multiply these all together we get:

$$11 \cdot 10 \cdot 9 \cdot \binom{8}{2} \cdot 6 \cdot \binom{5}{2} \cdot 3 \cdot 2 \cdot 1 = \frac{11!}{2! \cdot 2!}$$

(b) Here are two ways to do this problem.

Method 1. Since every arrangement has equal probability of being chosen we simply have to count the number that start with the letter 'b'. After putting a 'b' in position 1 there are 10 letters: p, r, o, b, a, i,i, l, t, y, to place in the last 10 positions. We count this in the same manner as part (a). That is

Choose the position for p: 10 ways.

Choose the positions for r,o,b,a,: $9 \cdot 8 \cdot 7 \cdot 6$ ways.

Choose two positions for the two i's: 5 choose 2 ways.

Choose the position for l: 3 ways.

Choose the position for t: 2 ways.

Choose the position for y: 1 ways.

Multiplying this together we get $10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot \binom{5}{2} \cdot 3 \cdot 2 \cdot 1 = \frac{10!}{2!}$ arrangements start with the

letter b. Therefore the probability a random arrangement starts with b is $\frac{10!/2!}{11!/2! \cdot 2!} = \frac{2}{11}$

Method 2. Suppose we build the arrangement by picking a letter for the first position, then the second position etc. Since there are 11 letters, two of which are b's we have a $2/11$ chance of picking a b for the first letter.

Problem 2. (Probability)

We are given $P(E \cup F) = 3/4$.

$$E^c \cap F^c = (E \cup F)^c \Rightarrow P(E^c \cap F^c) = 1 - P(E \cup F) = \boxed{1/4}.$$

Problem 3. (Counting)

Let H_i be the event that the i^{th} hand has one king. We have the conditional probabilities

$$P(H_1) = \frac{\binom{4}{1} \binom{48}{12}}{\binom{52}{13}}; \quad P(H_2|H_1) = \frac{\binom{3}{1} \binom{36}{12}}{\binom{39}{13}}; \quad P(H_3|H_1 \cap H_2) = \frac{\binom{2}{1} \binom{24}{12}}{\binom{26}{13}}$$

$$P(H_4|H_1 \cap H_2 \cap H_3) = 1$$

$$\begin{aligned} P(H_1 \cap H_2 \cap H_3 \cap H_4) &= P(H_4|H_1 \cap H_2 \cap H_3) P(H_3|H_1 \cap H_2) P(H_2|H_1) P(H_1) \\ &= \frac{\binom{2}{1} \binom{24}{12} \binom{3}{1} \binom{36}{12} \binom{4}{1} \binom{48}{12}}{\binom{26}{13} \binom{39}{13} \binom{52}{13}}. \end{aligned}$$

Problem 4. (Conditional probability)

(a) Sample space = $\Omega = \{(1, 1), (1, 2), (1, 3), \dots, (6, 6)\} = \{(i, j) \mid i, j = 1, 2, 3, 4, 5, 6\}$.
(Each outcome is equally likely, with probability $1/36$.)

$$A = \{(1, 3), (2, 2), (3, 1)\},$$

$$B = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (1, 3), (2, 3), (4, 3), (5, 3), (6, 3)\}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/36}{11/36} = \frac{2}{11}.$$

(b) $P(A) = 3/36 \neq P(A|B)$, so they are not independent.

Problem 5. (Bayes formula)

For a given problem let C be the event the student gets the problem correct and K the event the student knows the answer.

The question asks for $P(K|C)$.

$$\text{We'll compute this using Bayes' rule: } P(K|C) = \frac{P(C|K) P(K)}{P(C)}.$$

$$\text{We're given: } P(C|K) = 1, \quad P(K) = 0.6.$$

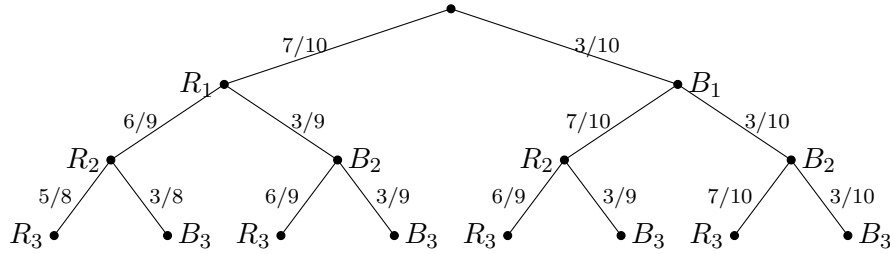
Law of total prob.:

$$P(C) = P(C|K) P(K) + P(C|K^c) P(K^c) = 1 \cdot 0.6 + 0.25 \cdot 0.4 = 0.7.$$

$$\text{Therefore } P(K|C) = \frac{0.6}{0.7} = .857 = 85.7\%.$$

Problem 6. (Bayes formula)

Here is the game tree, R_1 means red on the first draw etc.



Summing the probability to all the B_3 nodes we get

$$P(B_3) = \frac{7}{10} \cdot \frac{6}{9} \cdot \frac{3}{8} + \frac{7}{10} \cdot \frac{3}{9} \cdot \frac{3}{9} + \frac{3}{10} \cdot \frac{7}{10} \cdot \frac{3}{9} + \frac{3}{10} \cdot \frac{3}{10} \cdot \frac{3}{10} = .350.$$

Problem 7. (Expected value and variance)

| | | | | | |
|------|------|------|------|------|------|
| X | -2 | -1 | 0 | 1 | 2 |
| p(X) | 1/15 | 2/15 | 3/15 | 4/15 | 5/15 |

We compute

$$E[X] = -2 \cdot \frac{1}{15} + -1 \cdot \frac{2}{15} + 0 \cdot \frac{3}{15} + 1 \cdot \frac{4}{15} + 2 \cdot \frac{5}{15} = \frac{2}{3}.$$

Thus

$$\text{Var}(X) = E((X - \frac{2}{3})^2) = \frac{14}{9}.$$

Problem 8. (Expected value and variance)

We first compute

$$E[X] = \int_0^1 x \cdot 2x dx = \frac{2}{3}$$

$$E[X^2] = \int_0^1 x^2 \cdot 2x dx = \frac{1}{2}$$

$$E[X^4] = \int_0^1 x^4 \cdot 2x dx = \frac{1}{3}.$$

Thus,

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}$$

and

$$\text{Var}(X^2) = E[X^4] - (E[X^2])^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Problem 9. (Expected value and variance)

Use $\text{Var}(X) = E(X^2) - E(X)^2 \Rightarrow 3 = E(X^2) - 4 \Rightarrow E(X^2) = 7.$

Problem 10. (Expected value and variance)

answer:

Make a table

| | | |
|-------|---------|-----|
| X : | 0 | 1 |
| prob: | $(1-p)$ | p |
| X^2 | 0 | 1. |

From the table, $E(X) = 0 \cdot (1-p) + 1 \cdot p = p$.

Since X and X^2 have the same table $E(X^2) = E(X) = p$.

Therefore, $\text{Var}(X) = p - p^2 = p(1-p)$.

Problem 11. (Expected value)

Let X be the number of people who get their own hat.

Following the hint: let X_j represent whether person j gets their own hat. That is, $X_j = 1$ if person j gets their hat and 0 if not.

We have, $X = \sum_{j=1}^{100} X_j$, so $E(X) = \sum_{j=1}^{100} E(X_j)$.

Since person j is equally likely to get any hat, we have $P(X_j = 1) = 1/100$. Thus, $X_j \sim \text{Bernoulli}(1/100) \Rightarrow E(X_j) = 1/100 \Rightarrow E(X) = 1$.

Problem 12. (Expected value and variance)

(a) There are a number of ways to present this.

$X \sim 3 \text{ binomial}(25, 1/6)$, so

$$P(X = 3k) = \binom{25}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{25-k}, \quad \text{for } k = 0, 1, 2, \dots, 25.$$

(b) $X \sim 3 \text{ binomial}(25, 1/6)$.

Recall that the mean and variance of $\text{binomial}(n, p)$ are np and $np(1-p)$. So,

$$E(X) = 3np = 3 \cdot 25 \cdot \frac{1}{6} = 75/6, \quad \text{and } \text{Var}(X) = 9np(1-p) = 9 \cdot 25(1/6)(5/6) = 125/4.$$

(c) $E(X + Y) = E(X) + E(Y) = 150/6 = 25$, $E(2X) = 2E(X) = 150/6 = 25$.

$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = 250/4$. $\text{Var}(2X) = 4\text{Var}(X) = 500/4$.

The means of $X + Y$ and $2X$ are the same, but $\text{Var}(2X) > \text{Var}(X + Y)$.

This makes sense because in $X + Y$ sometimes X and Y will be on opposite sides from the mean so distances to the mean will tend to cancel, However in $2X$ the distance to the mean is always doubled.

Problem 13. (Continuous random variables)

First we find the value of a :

$$\int_0^1 f(x) dx = 1 = \int_0^1 x + ax^2 dx = \frac{1}{2} + \frac{a}{3} \Rightarrow a = 3/2.$$

The CDF is $F_X(x) = P(X \leq x)$. We break this into cases:

(i) $b < 0 \Rightarrow F_X(b) = 0$.

(ii) $0 \leq b \leq 1 \Rightarrow F_X(b) = \int_0^b x + \frac{3}{2}x^2 dx = \frac{b^2}{2} + \frac{b^3}{2}$.

(iii) $1 < x \Rightarrow F_X(b) = 1$.

Using F_X we get

$$P(.5 < X < 1) = F_X(1) - F_X(.5) = 1 - \left(\frac{.5^2 + .5^3}{2} \right) = \frac{13}{16}.$$

Problem 14. (Exponential distribution)

(a) We compute

$$P(X \geq 5) = 1 - P(X < 5) = 1 - \int_0^5 \lambda e^{-\lambda x} dx = 1 - (1 - e^{-5\lambda}) = e^{-5\lambda}.$$

(b) We want $P(X \geq 15 | X \geq 10)$. First observe that $P(X \geq 15, X \geq 10) = P(X \geq 15)$. From similar computations in (a), we know

$$P(X \geq 15) = e^{-15\lambda} \qquad P(X \geq 10) = e^{-10\lambda}.$$

From the definition of conditional probability,

$$P(X \geq 15 | X \geq 10) = \frac{P(X \geq 15, X \geq 10)}{P(X \geq 10)} = \frac{P(X \geq 15)}{P(X \geq 10)} = e^{-5\lambda}$$

Note: This is an illustration of the memorylessness property of the exponential distribution.

Problem 15. (a) Note, Y follows what is called a *log-normal distribution*.

$$F_Y(a) = P(Y \leq a) = P(e^Z \leq a) = P(Z \leq \ln(a)) = \boxed{\Phi(\ln(a))}.$$

Differentiating using the chain rule:

$$f_y(a) = \frac{d}{da} F_Y(a) = \frac{d}{da} \Phi(\ln(a)) = \frac{1}{a} \phi(\ln(a)) = \boxed{\frac{1}{\sqrt{2\pi} a} e^{-(\ln(a))^2/2}}.$$

(b) (i) We want to find $q_{.33}$ such that $P(Z \leq q_{.33}) = .33$. That is, we want

$$\Phi(q_{.33}) = .33 \Leftrightarrow \boxed{q_{.33} = \Phi^{-1}(.33)}.$$

(ii) We want $q_{.9}$ such that

$$F_Y(q_{.9}) = .9 \Leftrightarrow \Phi(\ln(q_{.9})) = .9 \Leftrightarrow \boxed{q_{.9} = e^{\Phi^{-1}(.9)}}.$$

(iii) As in (ii) $q_{.5} = e^{\Phi^{-1}(.5)} = e^0 = \boxed{1}$.

Problem 16. (a) We did this in class. Let $\phi(z)$ and $\Phi(z)$ be the PDF and CDF of Z .
 $F_Y(y) = P(Y \leq y) = P(aZ + b \leq y) = P(Z \leq (y - b)/a) = \Phi((y - b)/a)$.

Differentiating:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \Phi((y - b)/a) = \frac{1}{a} \phi((y - b)/a) = \frac{1}{\sqrt{2\pi} a} e^{-(y-b)^2/2a^2}.$$

Since this is the density for $N(b, a^2)$ we have shown $Y \sim N(b, a^2)$.

(b) By part (a), $Y \sim N(\mu, \sigma^2) \Rightarrow Y = \sigma Z + \mu$.

But, this implies $(Y - \mu)/\sigma = Z \sim N(0, 1)$. QED

Problem 17. (Quantiles)

The density for this distribution is $f(x) = \lambda e^{-\lambda x}$. We know (or can compute) that the distribution function is $F(a) = 1 - e^{-\lambda a}$. The median is the value of a such that $F(a) = .5$. Thus, $1 - e^{-\lambda a} = 0.5 \Rightarrow 0.5 = e^{-\lambda a} \Rightarrow \log(0.5) = -\lambda a \Rightarrow \boxed{a = \log(2)/\lambda}$.

Problem 18. (Correlation)

As usual let X_i = the number of heads on the i^{th} flip, i.e. 0 or 1.

Let $X = X_1 + X_2 + X_3$ the sum of the first 3 flips and $Y = X_3 + X_4 + X_5$ the sum of the last 3. Using the algebraic properties of covariance we have

$$\begin{aligned} \text{Cov}(X, Y) &= \text{Cov}(X_1 + X_2 + X_3, X_3 + X_4 + X_5) \\ &= \text{Cov}(X_1, X_3) + \text{Cov}(X_1, X_4) + \text{Cov}(X_1, X_5) \\ &\quad + \text{Cov}(X_2, X_3) + \text{Cov}(X_2, X_4) + \text{Cov}(X_2, X_5) \\ &\quad + \text{Cov}(X_3, X_3) + \text{Cov}(X_3, X_4) + \text{Cov}(X_3, X_5) \end{aligned}$$

Because the X_i are independent the only non-zero term in the above sum is $\text{Cov}(X_3, X_3) = \text{Var}(X_3) = \frac{1}{4}$

Therefore, $\text{Cov}(X, Y) = \frac{1}{4}$.

We get the correlation by dividing by the standard deviations. Since X is the sum of 3 independent Bernoulli(.5) we have $\sigma_X = \sqrt{3/4}$

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{1/4}{(3)/4} = \frac{1}{3}.$$

Problem 19. (Joint distributions)

(a) Here we have two continuous random variables X and Y with joint probability density function

$$f(x, y) = \frac{12}{5} xy(1 + y) \quad \text{for } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1,$$

and $f(x, y) = 0$ otherwise. So

$$P\left(\frac{1}{4} \leq X \leq \frac{1}{2}, \frac{1}{3} \leq Y \leq \frac{2}{3}\right) = \int_{\frac{1}{4}}^{\frac{1}{2}} \int_{\frac{1}{3}}^{\frac{2}{3}} f(x, y) dy dx = \frac{41}{720}.$$

(b) $F(a, b) = \int_0^a \int_0^b f(x, y) dy dx = \frac{3}{5}a^2b^2 + \frac{2}{5}a^2b^3$ for $0 \leq a \leq 1$ and $0 \leq b \leq 1$.

(c) We find the marginal cdf $F_X(a)$ by setting b in $F(a, b)$ to the top of its range, i.e. $b = 1$.

So $F_X(a) = F(a, 1) = a^2$.

(d) For $0 \leq x \leq 1$, we have

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 f(x, y) dy = 2x.$$

This is consistent with (c) because $\frac{d}{dx}(x^2) = 2x$.

(e) We first compute $f_Y(y)$ for $0 \leq y \leq 1$ as

$$f_Y(y) = \int_0^1 f(x, y) dx = \frac{6}{5}y(y+1).$$

Since $f(x, y) = f_X(x)f_Y(y)$, we conclude that X and Y are independent.

Problem 20. (Joint distributions)

(a) The marginal probability $P_Y(1) = 1/2$

$\Rightarrow P(X = 0, Y = 1) = P(X = 2, Y = 1) = 0$.

Now each column has one empty entry. This can be computed by making the column add up to the given marginal probability.

| $Y \setminus X$ | 0 | 1 | 2 | P_Y |
|-----------------|-----|-----|-----|-------|
| -1 | 1/6 | 1/6 | 1/6 | 1/2 |
| 1 | 0 | 1/2 | 0 | 1/2 |
| P_X | 1/6 | 2/3 | 1/6 | 1 |

(b) No, X and Y are not independent.

For example, $P(X = 0, Y = 1) = 0 \neq 1/12 = P(X = 0) \cdot P(Y = 1)$.

Problem 21.

Standardize:

$$\begin{aligned} P\left(\sum_i X_i < 30\right) &= P\left(\frac{\frac{1}{n} \sum X_i - \mu}{\sigma/\sqrt{n}} < \frac{30/n - \mu}{\sigma/\sqrt{n}}\right) \\ &\approx P\left(Z < \frac{30/100 - 1/5}{1/30}\right) \quad (\text{by the central limit theorem}) \\ &= P(Z < 3) \\ &= 0.9987 \quad (\text{from the table of normal probabilities}) \end{aligned}$$

Problem 22. (Central limit theorem)

Let X_j be the IQ of a randomly selected person. We are given $E(X_j) = 100$ and $\sigma_{X_j} = 15$.

Let \bar{X} be the average of the IQ's of 100 randomly selected people. We have $E(\bar{X}) = 100$ and $\sigma_{\bar{X}} = 15/\sqrt{100} = 1.5$.

The problem asks for $P(\bar{X} > 115)$. Standardizing we get $P(\bar{X} > 115) \approx P(Z > 10)$.

This is effectively 0.

Problem 23. (NHST chi-square)

We will use a chi-square test for homogeneity. Remember we need to [use all the data!](#). For hypotheses we have:

H_0 : the re-offense rate is the same for both groups.

H_A : the rates are different.

Here is the table of counts. The computation of the expected counts is explained below.

| | Control group | | Experimental group | | |
|-----------------|---------------|----------|--------------------|----------|-----|
| | observed | expected | observed | expected | |
| Re-offend | 70 | 50 | 30 | 50 | 100 |
| Don't re-offend | 130 | 150 | 170 | 150 | 300 |
| | 200 | | 200 | | 400 |

The expected counts are computed as follows. Under H_0 the re-offense rates are the same, say θ . To find the expected counts we find the MLE of θ using the combined data:

$$\hat{\theta} = \frac{\text{total re-offend}}{\text{total subjects}} = \frac{100}{400}.$$

Then, for example, the expected number of re-offenders in the control group is $200 \cdot \hat{\theta} = 50$. The other expected counts are computed in the same way.

The chi-square test statistic is

$$X^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = \frac{20^2}{50} + \frac{20^2}{150} + \frac{20^2}{50} + \frac{20^2}{150} \approx 8 + 2.67 + 8 + 2.67 \approx 21.33.$$

Finally, we need the degrees of freedom: $df = 1$ because this is a two-by-two table and $(2 - 1) \cdot (2 - 1) = 1$. (Or because we can freely fill in the count in one cell and still be consistent with the marginal counts 200, 200, 100, 300, 400 used to compute the expected counts.)

From the χ^2 table: $p = P(X^2 > 21.33 | df = 1) < 0.01$.

Conclusion: we reject H_0 in favor of H_A . The experimental intervention appears to be effective.

Problem 24. (Confidence intervals)

We compute the data mean and variance $\bar{x} = 65$, $s^2 = 35.778$. The number of degrees of freedom is 9. We look up the *critical value* $t_{9,0.025} = 2.262$ in the t -table. The 95% confidence interval is

$$\left[\bar{x} - \frac{t_{9,0.025}s}{\sqrt{n}}, \bar{x} + \frac{t_{9,0.025}s}{\sqrt{n}} \right] = \left[65 - 2.262\sqrt{3.5778}, 65 + 2.262\sqrt{3.5778} \right] = [60.721, 69.279]$$

Problem 25. (Confidence intervals)

Suppose we have taken data x_1, \dots, x_n with mean \bar{x} . The 95% confidence interval for the mean is $\bar{x} \pm z_{0.025} \frac{\sigma}{\sqrt{n}}$. This has width $2 z_{0.025} \frac{\sigma}{\sqrt{n}}$. Setting the width equal to 1 and substituting values $z_{0.025} = 1.96$ and $\sigma = 5$ we get

$$2 \cdot 1.96 \frac{5}{\sqrt{n}} = 1 \Rightarrow \sqrt{n} = 19.6.$$

So, $n = (19.6)^2 = \boxed{384}$.

If we use our rule of thumb that $z_{0.025} = 2$ we have $\sqrt{n}/10 = 2 \Rightarrow n = 400$.

Problem 26. (Confidence intervals)

The 90% confidence interval is $\bar{x} \pm z_{0.05} \cdot \frac{1}{2\sqrt{n}}$. Since $z_{0.05} = 1.64$ and $n = 400$ our confidence interval is

$$\bar{x} \pm 1.64 \cdot \frac{1}{40} = \bar{x} \pm 0.041$$

If this is entirely above 0.5 we have $\bar{x} - 0.041 > 0.5$, so $\bar{x} > 0.541$. Let T be the number out of 400 who prefer A. We have $\bar{x} = \frac{T}{400} > 0.541$, so $\boxed{T > 216}$.

Problem 27. (Confidence intervals)

A 95% confidence means about 5% = 1/20 will be wrong. You'd expect about 2 to be wrong.

With a probability $p = 0.05$ of being wrong, the number wrong follows a Binomial(40, p) distribution. This has expected value 2, and standard deviation $\sqrt{40(0.05)(0.95)} = 1.38$. 10 wrong is $(10-2)/1.38 = 5.8$ standard deviations from the mean. This would be surprising.

Problem 28. (Confidence intervals)

We have $n = 20$ and $s^2 = 4.06^2$. If we fix a hypothesis for σ^2 we know

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

We used R to find the critical values. (Or use the χ^2 table.)

`c025 = qchisq(0.975, 19) = 32.852`

`c975 = qchisq(0.025, 19) = 8.907`

The 95% confidence interval for σ^2 is

$$\left[\frac{(n-1) \cdot s^2}{c_{0.025}}, \frac{(n-1) \cdot s^2}{c_{0.975}} \right] = \left[\frac{19 \cdot 4.06^2}{32.852}, \frac{19 \cdot 4.66^2}{8.907} \right] = [9.53, 35.16]$$

We can take square roots to find the 95% confidence interval for σ

$$[3.09, 5.93]$$

Problem 29. (a) Step 1. We have the point estimate $p \approx \hat{p} = 0.30303$.

Step 2. Use the computer to generate many (say 10000) size 100 samples. (These are called the bootstrap samples.)

Step 3. For each sample compute $p^* = 1/\bar{x}^*$ and $\delta^* = p^* - \hat{p}$.

Step 4. Sort the δ^* and find the critical values $\delta_{0.95}$ and $\delta_{0.05}$. (Remember $\delta_{0.95}$ is the 5th percentile etc.)

Step 5. The 90% bootstrap confidence interval for p is

$$[\hat{p} - \delta_{0.05}, \hat{p} - \delta_{0.95}]$$

(b) It's tricky to keep the sides straight here. We work slowly and carefully:
The 5th and 95th percentiles for \bar{x}^* are the 10th and 190th entries

$$2.89, \quad 3.72$$

(Here again there is some ambiguity on which entries to use. We will accept using the 11th or the 191st entries or some interpolation between these entries.)

So the 5th and 95th percentiles for p^* are

$$1/3.72 = 0.26882, \quad 1/2.89 = 0.34602$$

So the 5th and 95th percentiles for $\delta^* = p^* - \hat{p}$ are

$$-0.034213, \quad 0.042990$$

These are also the 0.95 and 0.05 critical values.

So the 90% CI for p is

$$[0.30303 - 0.042990, 0.30303 + 0.034213] = [0.26004, 0.33724]$$

Problem 30. (a) The steps are the same as in the previous problem except the bootstrap samples are generated in different ways.

Step 1. We have the point estimate $q_{0.5} \approx \hat{q}_{0.5} = 3.3$.

Step 2. Use the computer to generate many (say 10000) size 100 resamples of the original data.

Step 3. For each sample compute the median $q_{0.5}^*$ and $\delta^* = q_{0.5}^* - \hat{q}_{0.5}$.

Step 4. Sort the δ^* and find the critical values $\delta_{0.95}$ and $\delta_{0.05}$. (Remember $\delta_{0.95}$ is the 5th percentile etc.)

Step 5. The 90% bootstrap confidence interval for $q_{0.5}$ is

$$[\hat{q}_{0.5} - \delta_{0.05}, \hat{q}_{0.5} - \delta_{0.95}]$$

(b) This is very similar to the previous problem. We proceed slowly and carefully to get terms on the correct side of the inequalities.

The 5th and 95th percentiles for $q_{0.5}^*$ are

$$2.89, \quad 3.72$$

So the 5th and 95th percentiles for $\delta^* = q_{0.5}^* - \hat{q}_{0.5}$ are

$$[2.89 - 3.3, 3.72 - 3.3] = [-0.41, 0.42]$$

These are also the 0.95 and 0.05 critical values.

So the 90% CI for p is

$$[3.3 - 0.42, 3.3 + 0.41] = [2.91, 3.71]$$

Problem 31. The model is $y_i = a + bx_i + \varepsilon_i$, where ε_i is random error. We assume the errors are independent with mean 0 and the same variance for each i (homoscedastic).

The total error squared is

$$E^2 = \sum (y_i - a - bx_i)^2 = (1 - a - b)^2 + (1 - a - 2b)^2 + (3 - a - 3b)^2$$

The least squares fit is given by the values of a and b which minimize E^2 . We solve for them by setting the partial derivatives of E^2 with respect to a and b to 0. In R we found that $a = 1.0$, $b = 0.5$

MIT OpenCourseWare
<https://ocw.mit.edu>

18.05 Introduction to Probability and Statistics
Spring 2014

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.