Math 3070 § 1. $\sigma\tau$ **Third Midquarter Sample Exam Solutions**
Treibergs November 7, 2001

(1.) *Suppose that in a random sample of 26 dentists surveyed, 19 recommend the use of "Snow-White" toothpaste. Does this justify with 95% confidence that "at least 7 of 10 all dentists recommend the use of Snow-White toothpaste"? At least what proportion can you be 95% confident are recommending this product?*

We wish to construct one sided lower confidence bounds the proportion of dentists that recommend SW. Thus the critical z-value that cuts off one tail with 0.05 area is $z = z_{.05} = 1.645$. Our estimator of proportion is $\hat{p} = 19/26 = .731$ and sample size $n = 26$. Since $\hat{p}n \geq \hat{q}n = (.269) \cdot 26 = 6.995 \geq 10$ it follows we can use the large sample confidence interval for proportion. Thus, with 95% confidence,

$$.588 = .731 - (1.645)\sqrt{\frac{(.731)(.269)}{26}} = \hat{p} - z_\alpha \sqrt{\frac{\hat{p}\hat{q}}{n}} < p$$

A confidence bound that works better (is more accurate and is not restricted by the rule of thumb) than the large sample bound in the box on p. 265 may be obtained by solving the inequality $P((\hat{p} - p)/\sqrt{p(1-p)/n} < z_\alpha) = 1 - \alpha$ for p in the second equation of that page. It amounts to resolving the quadratic inequality to get an unrestricted one sided formula formula (use $\pm z_{\alpha/2}$ in place of z_α for the two-sided bounds!)

$$\frac{\hat{p} + \frac{z_\alpha^2}{2n} - z_\alpha \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_\alpha^2}{4n^2}}}{1 + z_\alpha^2/n} \leq p$$

With $1 - \alpha = .95$ confidence, the actual proportion satisfies

$$.571 = \frac{(.731) + \frac{(1.645)^2}{2 \cdot 26} - (1.645)\sqrt{\frac{(.731)(.269)}{26} + \frac{(1.645)^2}{4(26)^2}}}{1 + (1.645)^2/26} \leq p$$

Thus with 95% confidence, $0.7 \leq p$ is not justified from this limited data, but $.571 \leq p$ is. The sharper bounds don't need to satisfy the restrictions $\hat{p} \geq 5$ and $\hat{q} \geq 5$ which are the rule of thumb for when the large sample bounds work.

(2.) *Suppose the following annual rainfall readings in inches were observed at Dead Horse Point for the years 1992 to 2001. Construct a normal quantile-quantile plot for these readings. Is normality plausible?*

6.47, 8.07, 8.34, 9.21, 3.54, 3.39, 10.91, 6.54, 4.61, 5.51

Let $\Phi(z) = P(Z \leq z)$ where Z is a standard normal variable. The normal probability plot is obtained as follows. We sort the data so $y_1 = 3.39, \ldots, y_{10} = 10.91$. We find the critical normal values $\Phi(x_i) = (i - 0.375)/(n - .250) = f$ so that $x_1 = z_{.936} \approx 4.91(f^{0.14} + (1-f)^{0.14}) = -1.522$ and so on. Then plot the points $[x_i, y_i]$. Here is the complete list: $[-1.522, 3.39]$, $[-.966, 3.54]$, $[-.613, 4.61]$, $[-.326, 5.51]$, $[-.064, 6.47]$, $[.064, 6.54]$, $[.326, 8.07]$, $[.613, 8.34]$, $[.966, 9.21]$, $[1.522, 10.91]$.

The data lines up quite well so normality of the data is plausible. (In fact a normal random number generator generated the data.)

(3.) *From the rainfall readings in problem (2.), estimate the mean of the annual rainfall. Estimate the standard error of your point estimate. Assuming that the distribution is normal, find a 90% two sided confidence interval for the mean rainfall.*

We compute the sample mean $\overline{X} = 6.659$, which is the estimator for the mean rainfall μ, and sample standard deviation to be $S = 2.353$. The standard error is the standard deviation of the statistic $\sigma_{\overline{X}} = \sigma/\sqrt{n}$ which we approximate to be $\hat{\sigma}_{\overline{X}} = S/\sqrt{n} = 2.353/\sqrt{10} = .744$. As the sample size is small $n \le 30$ we use Student's t-distribution. The degrees of freedom $\nu = n-1 = 9$ and since we want two sided bounds at the $1-\alpha = .90$ confidence level, each tail is has .05 area and the critical t-value is $t_{.05,9} = 1.833$. Thus the two sided confidence intervals are

$$5.30 = 6.659 - (1.833)(.744) = \overline{X} - \frac{t_{.05,9}S}{\sqrt{n}} \le \mu \le \overline{X} + \frac{t_{.05,9}S}{\sqrt{n}} = 6.659 + (1.833)(.744) = 8.02.$$

(4.) *For the same data as in Problem(2.) With 95% confidence, find tolerance bounds for the amount of rainfall that Dead Horse Point will see 90 years out of the next 100?*

$\overline{X}$, S and ν as before. The .95 tolerance interval is found in table A.7. We want a $1-\gamma = 95\%$ confidence, two sided tolerance bound that captures $1-\alpha = 90\%$ of the population, which is $\mathrm{tol}_{1-\gamma=.95, 1-\alpha=.90, n=10} = 2.839$. Thus, using the tolerance interval on p. 250, we may be 95% confident that 90 years of 100, the rainfall will satisfy

$$\begin{aligned}-.021 = 6.659 - (2.839)(2.353) = \overline{X} - \mathrm{k}_{\alpha=.95,k=.90,n=10}\, S \le X \\ \le \overline{X} + \mathrm{k}\, S = 6.659 - (2.839)(2.353) = 13.3.\end{aligned}$$

(5.) *Let X be a continuous random variable with probability distribution $f(x) = (0.5) + (0.5)x$ if $-1 \le x \le 1$ and $f(x) = 0$ if $|x| > 1$. Find the probability distribution of the random variable $Y = X^4$.*

In this problem, there are two values of X that correspond to a given Y, namely $X = \pm Y^{1/4} = W(Y)$. Thus for $Y > 1$, $|X| = Y^{1/4} > 1$ so the distribution satisfies $g(y) = f(x) = 0$. There are two disjoint sets, namely $-1 \le x < 0$ and $0 < x \le 1$ on which x^4 is monotone and map onto $(0,1]$. Thus $w_1(y) = -y^{1/4}$ corresponds to $[-1, 0)$ and $w_2(y) = y^{1/4}$ correspods to $(0,1]$. Thus $w_1' = -\frac{1}{4}y^{-3/4}$ and $w_2'(y) = \frac{1}{4}y^{-1/4}$. According to Theorem 7.5, the probability distribution of Y is then

$$g(y) = \begin{cases} f(w_1(y))|w_1'(y)| + f(w_2(y))|w_2'(y)| = \frac{y^{-3/4} - y^{-1/2}}{8} + \frac{y^{-3/4} + y^{-1/2}}{8} = \frac{y^{-3/4}}{4}, & \text{if } 0 < y \le 1; \\ 0, & \text{otherwise.} \end{cases}$$

Note that $P(\varepsilon < y \le 1) \to 1$ as $\varepsilon \to 0$ so there is no loss by putting $g(0) = 0$.

(6.) *An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with standard deviation of 40 hrs. If a random sample of 45 bulbs has an average of 780 hours, find a 96% confidence interval for the population mean of all bulbs produced by this firm.*

Since the distribution is approximately normal, and $n > 30$ the standardized sample mean behaves almost normally, and we compute the two sided confidence interval using the z-critical value. 96% leaves 2% for each tail, thus $\Phi(z_{.02}) = .9800$ falls between $\Phi(2.05) = .9798$ and $\Phi(2.06) = .9803$. Interpolating, $z_{.03} = 2.054$. Or use table A4 with $\nu = \infty$ d.f. Thus with 96% confidence,

$$768. = 780. - \frac{40.0 \cdot 2.054}{\sqrt{45}} = \overline{X} - \frac{S z_{.02}}{\sqrt{n}} \le \mu \le \overline{X} + \frac{S z_{.02}}{\sqrt{n}} = 780. - \frac{40.0 \cdot 2.054}{\sqrt{45}} = 792.$$

(7.) *Suppose three airlines, $1, 2, 3$ fly from Salt Lake City to San Francisco. Suppose that the fares (in \$) are independent, normally distributed with means and standard deviations $\mu_1 = 300$, $\mu_2 = 350$, $\mu_3 = 275$, $\sigma_1 = 15$, $\sigma_2 = 25$, $\sigma_1 = 35$. What is the probability that the cost of two such flights on airline 1 will exceed the sum of the costs of one flight on airline 2 and one on airline 3?*

Setting the random variable $Y = -2X_1 + X2 + X3$ the question is asking for $P(2X_1 \geq X_2 + X_3) = P(Y \leq 0)$. The expectation is $\mu_Y = -2\mu_1 + \mu_2 + \mu_3 = -2 \cdot 300. + 350. + 275. = 25.$ Since the rv's are assumed to be independent, the variance satisfies

$$\sigma_Y^2 = (-2)^2\sigma_1^2 + (1)^2\sigma_2^2 + (1)^2\sigma_3^2 = 4 \cdot (15)^2 + (25)^2 + (35)^2 = 2750.$$

Then by standardizing to z-scores, ($Y_0 = 0$,)

$$P(Y \leq 0) = P\left(Z \leq \tfrac{Y_0 - \mu_Y}{\sigma_Y}\right) = P\left(Z \leq \tfrac{0-25.}{52.44}\right) = \Phi(-.477) = .3168.$$

(8.) *Suppose and $X_1 \in \{0, 3\}$ and $X_2 \in \{0, 2, 4\}$ are discrete random variables with the given joint pmf. Find the joint probability distribution of $Y_1 = X_1 + X_2$ and $Y_2 = X_1X_2$. Find the marginal distribution of Y_2.*

		$x_2 = 0$	2	4
pmf(x_1,x_2)	$x_1 = 0$	0.2	0.3	0.08
	3	0.15	0.15	0.12

The table values $(x_1, y_2) \in \left\{\begin{matrix} (0,0), & (0,2), & (0,4), \\ (3,0), & (3,2), & (3,4) \end{matrix}\right\}$ correspond term by term to $y_1 = x_1 + x_2$, $y_2 = x_1x_2$

$(y_1, y_2) \in \left\{\begin{matrix} (0,0), & (2,0), & (4,0), \\ (3,0), & (5,6), & (7,12) \end{matrix}\right\}$ thus $g(y_1, y_2) = f(u(y_1, y_2), v(y{-}1, y_2))$ where $x_1 = u(y_1, y_2)$ and $x_2 = v(y_1, y_2)$. Thus $g(0, 0) = f(0, 0) = .2$, $g(2, 0) = f(0, 2) = .3$, $g(4, 0) = f(0, 4) = .08$, $g(3, 0) = f(3, 0) = .15$, $g(5, 6) = f(3, 2) = .15$, $g(7, 12) = f(3, 6) = .12$ and $g(y_1, y_2) = 0$ for any other pair (y_1, y_2). The values taken by $Y_2 \in \{0, 6, 12\}$. Thus the marginal probability for Y_2 is $h(0) = g(0, 0) + g(2, 0) + g(3, 0) + g(4, 0) = .2 + .3 + .15 + .08 = .73$, $h(6) = g(5, 6) = .15$, $h(12) = g(7, 12) = .12$ and $h(y_2) = 0$ otherwise.

(9.) *If S_1^2 and S_2^2 represent the variances of independent random samples of size $n_1 = 8$ and $n_2 = 13$, taken from normal populations with equal variances, find the $P(S_1^2/S_2^2 < 4.81)$.*

(10.) *Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Show that the sample variance is an unbiased estimator for the population variance σ^2.*

$$S^2 = \tfrac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2 = \tfrac{1}{n-1}\left\{\sum_{i=1}^n {X_i}^2 - \tfrac{1}{n}\left(\sum_{i=1}^n X_i\right)^2\right\}$$

A statistic is an *unbiased* estimator for σ^2 if its expectation is $E(S^2) = \sigma^2$. We are assuming independent variables which means that the formula for the variance of a sum simplifies. To see this, consider the case of X and Y beind independent continuous variables. That means that the joint distribution function can be factored $F(x, y) = f(x)f(y)$. Computing $E(XY)$ and the variance of the sum, (where $\int$ means $\int_{-\infty}^{\infty}$.)

$$\begin{aligned}
E(XY) &= \iint xyf(x)f(y)\,dxdy = \int yf(y)\left(\int xf(x)\,dx\right)\,dy \\
&= \int yf(y)E(X)\,dy = E(X)E(Y), \\
E((X+Y)^2) &= E(X^2+2XY+Y^2) = E(X^2)+2E(XY)+E(Y^2) \\
&= E(X^2)+2E(X)E(Y)+E(Y^2) \\
V(X+Y) &= E((X+Y)^2)-[E(X+Y)]^2 \\
&= E(X^2)+2E(X)E(Y)+E(Y^2)-[E(X)+E(Y)]^2 \\
&= E(X^2)+2E(X)E(Y)+E(Y^2)-[E(X)]^2-2E(X)E(Y)-[E(Y)]^2 \\
&= E(X^2)-[E(X)]^2+E(Y^2)-[E(Y)]^2 = V(X)+V(Y).
\end{aligned}$$

Thus for independent random variables, the variance of the sum is the sum of the variances. This also applies to sums of several independent random variable. Writing $\sum$ for $\sum_{i=1}^{n}$ and using $E(X^2) = V(X) + [E(X)]^2$, and that all the variables have the same means and variances, $\mu = E(X_i)$ and $\sigma^2 = V(X_i)$,

$$\begin{aligned}
E(S^2) =& \tfrac{1}{n-1}E\left(\Sigma X_i{}^2 - \tfrac{1}{n}\left[\Sigma X_i\right]^2\right) = \tfrac{1}{n-1}\Sigma E(X_i{}^2) - \tfrac{1}{n(n-1)}E\left(\left[\Sigma X_i\right]^2\right) \\
=& \tfrac{1}{n-1}\Sigma\left\{V(X_i)+[E(X_i)]^2\right\} - \tfrac{1}{n(n-1)}\left\{V\left(\Sigma X_i\right)+\left[E\left(\Sigma X_i\right)\right]^2\right\} \\
=& \tfrac{1}{n-1}\Sigma\left\{V(X_i)+[E(X_i)]^2\right\} - \tfrac{1}{n(n-1)}\left\{\Sigma V(X_i)+[\Sigma E(X_i)]^2\right\} \\
=& \tfrac{n}{n-1}\left\{\sigma^2+\mu^2\right\} - \tfrac{1}{n(n-1)}\left\{n\sigma^2+[n\mu]^2\right\} = \sigma^2.
\end{aligned}$$

(11.) *Suppose X is a continuous random variable which is uniformly distributed in $[-c, c]$. Find an unbiased point estimator for c. Hint: consider the statistic $\mathcal{M} = \max\{X, Y\}$.*

To find an estimator for c, we need to find some statistic whose expectation involves c. To compute the expectation of $\mathcal{M}$ we assume both X and Y are uniformly distributed in $[-c, c]$ and are independent so that the joint distribution function is the product of the marginal pdf's

$$f(x,y) = \begin{cases} \frac{1}{4c^2}, & \text{if } -c \le x, y \le c, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, since $x \ge y$ on the triangle $-c \le y \le x \le c$, and by symmetry $f(x,y) = f(y,x)$,

$$\begin{aligned}
E(\mathcal{M}) &= \int_{y=-\infty}^{\infty}\int_{x=-\infty}^{\infty} \max\{x,y\}f(x,y)\,dxdy \\
&= 2\int_{x=-c}^{c}\int_{y=-c}^{x}\frac{xdy\,dx}{4c^2} = \frac{1}{2c^2}\int_{x=-c}^{c}\Big[xy\Big]_{y=-c}^{x}dx \\
&= \frac{1}{2c^2}\int_{x=-c}^{c}x^2+cx\,dx = \frac{1}{2c^2}\left[\frac{x^3}{3}+\frac{cx^2}{2}\right]_{x=-c}^{c} \\
&= \frac{1}{2c^2}\left[\frac{c^3-(-c)^3}{3}+\frac{c(c^2-(-c)^2)}{2}\right]_{x=-c}^{c} = \frac{c}{3}.
\end{aligned}$$

Thus, an unbiased estimator for c is $\Theta = 3\mathcal{M} = 3\max\{X_1, X_2\}$.

Other estimators, such as the variance could have been used also.

(12.) *Let X be the number of pages in a randomly chosen airport novel. Although X can only assume positive integer values, it is approximately normally distributed with expected value 125 and standard deviation 16. What is the probability that a randomly chosen novel contains between 100 and 200 pages, using the continuity correction?*

(13.) *In a random sample of 900 homes in Boise, it is found that 202 are heated by oil. Find the 99% confidence interval for the proportion of homes in Boise heated by oil. How large a sample is needed to be 99% confident that our sample proportion will be within 0.05 of the true proportion of homes in Boise heated by oil?*

(14.) *An experiment was conducted to access the effect of using magnets at the filler point in the manufacture of coffee filter packs. We list the observed weights of filter packs in grams. Thirty packs produced with magnets are to be compared with 45 produced without magnets. Assuming that the population variances are equal, compute a two-sided 95% confidence interval for the difference of average weights of the filter packs made with and without magnets. What other assumptions are being made? [from Levine, Ramsey & Smidt "Applied Statistics," p. 420.]*

With Magnets

```
20.1 20.1 19.6 19.5 19.4 20.4 19.6 13.5 19.7 20.6 20.2 20.4 19.9 20.4 20.1
19.7 19.2 19.3 19.5 19.1 20.4 20.6 20.6 20.1 20.2 20.0 19.8 20.8 19.6 19.3
```

Without Magnets

```
21.4 20.5 20.2 20.6 21.4 20.1 19.9 19.7 21.2 20.7 20.9 20.2 20.1 20.7 20.4
21.0 20.5 20.0 20.6 20.8 19.5 19.7 20.0 20.4 20.2 20.5 20.4 21.3 20.4 20.2
20.5 20.7 19.8 20.7 20.2 20.0 20.5 20.6 20.6 19.9 20.0 20.2 20.6 19.9 20.4
```