Math 3070 § 1. **Third Midterm Exam** Name: Solutions
Treibergs $\sigma\tau$ April 2, 2008

(1.) *The molarity of a solute in solution is defined to be the number of moles per liter of solution (1 mole= 6.02×10^{23} molecules). If X is the molarity of solution of sodium chloride (NaCl) and Y is the molarity of a solution of sodium carbonate (Na_2CO_3), the molarity of the sodium ion (Na^+) in a solution made of equal parts of NaCl and Na_2CO_3 is given by $M = 0.5X + Y$. Assume X and Y are independent and normally distributed, and that X has mean 0.448 and standard deviation of 0.050, and Y has a mean of 0.253 and a standard deviation of 0.030. What is the distribution of M? Find $P(M > 0.5)$.*

Any linear combination of independent, normally distributed variables is also normally distributed $M \sim \mathrm{N}(\mu_M, \sigma_M)$ where $\mu_M = E(M) = E(0.5X+Y) = 0.5E(X)+E(Y) = 0.5\mu_X+\mu_Y = (0.5)(0.448) + 0.253 = \boxed{0.477}$. Since X and Y are independent, $\sigma_M^2 = V(M) = V(0.5X + Y) = (0.5)^2V(X) + (1)^2V(Y) = 0.25\sigma_X^2 + \sigma_Y^2 = (0.25)(0.050)^2 + (0.030)^2 = 0.001525$ so $\sigma_M = \boxed{0.0391}$. Thus standardizing, the probability $P(M > 0.5) = P\left(z > \frac{0.5-\mu_M}{\sigma_M}\right) = P\left(z > \frac{0.5-.477}{0.0391}\right) = P(z > .589) = P(z < -.589) = \boxed{.2779}$. Since interpolating $P(-.59) = .2776$ and $P(-.58) = .2810$ gives $P(-.59 + .001) \approx .2776 + (0.1)(.2810 - .2776) = .2779$.

(2.) *The Beams of Blanding Company has measured the nominal shear strength (in kN) for a random sample of 14 prestressed concrete beams. An analysis of the data found $\bar{X} = 696.8571$ and $s = 172.7643$. Find a 99% one-sided lower confidence interval for the shear strength. What assumptions are you making about the data that makes your test appropriate?*

Computing with $\nu = n - 1 = 14 - 1 = 13$ degrees of freedom and $99\% = (1 - \alpha) \times 100\%$ so $\alpha = .01$, we find one-tailed critical value $t_{\nu,\alpha} = t_{13,0.01} = 2.650$. Thus the lower 99% confidence bound is $\bar{X} - t_{\nu,\alpha}s/\sqrt{n} = 696.8571 - (2.650)(172.7643)/\sqrt{14} = \boxed{574.5}$. Since $n \le 40$ the large sample normal interval does not apply. σ is not known so the z-interval does not apply. The t-distribution based confidence interval is valid under the assumption that the random sample is taken from a normal distribution.

(3.) *Denote the triangle in the plane by $A = \{(x, y) : 0 \le x \le 1$ and $0 \le y \le 2x\}$. Suppose that the joint probability density function for a random point $(X, Y) \in \mathbb{R}^2$ is given by*

$$f(x,y) = \begin{cases} 1, & \text{if } (x,y) \in A, \\ 0, & \text{if } (x,y) \notin A. \end{cases}$$

Find the marginal densities of X and Y and the means μ_X and μ_Y. Find the covariance $Cov(X, Y)$. Are X and Y independent? Why?

The marginal densities are computed by integrating out the other variable. Thus $f_X(x) = 0$ if $x < 0$ or $x > 1$ and if $0 \le x \le 1$ then $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^{2x} dy = 2x$. Also $f_Y(y) = 0$ if $y < 0$ or $y > 2$ and if $0 \le y \le 2$ then $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_{y/2}^{1} dx = 1 - \frac{y}{2}$. Hence the means are $\mu_X = \int_{-\infty}^{\infty} x\, f_X(x)\,dx = \int_0^1 2x^2 = \frac{2}{3}$ and $\mu_Y = \int_{-\infty}^{\infty} y\, f_Y(y)\,dy = \int_0^2 y(1 - \frac{y}{2})\,dy = \frac{2}{3}$.

$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)\,dx\,dy = \int_0^1\int_0^{2x} xy\,dy\,dx = \frac{1}{2}$. Thus $\mathrm{Cov}(X,Y) = E(XY) - \mu_x\mu_y = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{18}$. X and Y are not independent because, if they were, their covariance would vanish. Another way to see it is that $f(x,y) \ne f_X(x)f_Y(y)$. For example, at the point $(x,y) = (0.1, 1)$, $f(0.1, 1) = 0$ but $f_X(0.1) = 0.2$ and $f_Y(1) = 0.5$.

(4.) *Let X, Y be independent random variables that are uniformly distributed on the interval $[-1 + c, 1 + c]$, where c is unknown. Show that both $\hat{\Theta}_1 = 0.4X + 0.6Y$ and $\hat{\Theta}_2 = 0.3X + 0.7Y$ are unbiased estimators for c. Which of the two is the better estimator and why?*

Both estimators are of the form $kX + (1-k)Y$ for some number k ($k = 0.4$ for $\hat{\Theta}_1$ and $k = 0.3$ for $\hat{\Theta}_2$). The density function $f(x; c) = \frac{1}{2}$ if $-1 + c \le x \le 1 + c$ and $f(x; c) = 0$ otherwise. Thus $E(X) = \int_{-\infty}^{\infty} x\, f(x;c)\,dx = \int_{-1+c}^{1+c} \frac{x}{2}dx = \left[\frac{x^2}{4}\right]_{-1+c}^{1+c} = \frac{1}{4}\left((1+c)^2 - (-1+c)^2\right) = c$. Thus the expectation of the estimator is $E(kX+(1-k)Y)=kE(X)+(1-k)E(Y)=kc+(1-k)c=c$ as both X and Y are taken from the same distribution. Since the expectation equals the parameter to be estimated, the statistics are *unbiased.*

Let us compare the sample variances. Since the width of the interval $[-1+c, 1+c]$ is independent of c, the variance $\sigma_X^2=\frac{1}{3}$ is independent of c also. Since X and Y are independent and identically distributed in a random sample, $V(\hat{\Theta}_1)=V(0.4X+0.6Y)=(0.4)^2V(X)+(0.6)^2V(Y)=(0.4)^2\sigma_X^2+(0.6)^2\sigma_Y^2=((0.4)^2+(0.6)^2)\sigma_X^2=0.52\sigma_X^2$ whereas $V(\hat{\Theta}_2)=V(0.3X+0.7Y)=(0.3)^2V(X)+(0.7)^2V(Y)=(0.3)^2\sigma_X^2+(0.7)^2\sigma_Y^2=((0.3)^2+(0.7)^2)\sigma_X^2=0.58\sigma_X^2$. Thus, using the principle that the lower the sample variance for the same number of observations, the better the statistic, we see that $\hat{\Theta}_1$ is the better estimator for c.

(5.) *A study to evaluate a method of averaging GPS readings to measure the altitude above sea level found that "large errors" were made at 9 of 24 sample test locations. Find a 90% two-sided confidence bound for the proportion of locations where the method made large errors. How many many measurements are needed to guarantee that the width of the confidence interval is no wider that 0.04 ?*

The estimator is $\hat{p}=\frac{9}{24}=\frac{3}{8}$ for $n=24$ observations so $\hat{q}=1-\hat{p}=\frac{5}{8}$. Since $n\hat{p}$ is less than 10, we must use the unrestricted CI for proportion. Since the confidence level is $90\%=(1-\alpha)\times 100\%$ we see that $\alpha=0.1$ and the two-sided CI depends on the critical value for the normal distribution $z_{\alpha/2}=z_{0.05}=1.645$. The two confidence bounds are given by the formula

$$\frac{\hat{p}+\frac{z_{\alpha/2}^2}{2n}\pm z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}+\frac{z_{\alpha/2}^2}{4n^2}}}{1+\frac{z_{\alpha/2}^2}{n}}=\frac{\frac{3}{8}+\frac{(1.645)^2}{2\cdot 24}\pm 1.645\sqrt{\frac{\frac{3}{8}\cdot\frac{5}{8}}{24}+\frac{(1.645)^2}{4(24)^2}}}{1+\frac{(1.645)^2}{24}}=0.388\pm 0.155.$$

For the width of the confidence interval to be $w=0.04$, and using the fact that $4\hat{p}\hat{q}\le 1$ we have that the estimate for the number

$$n\ge\frac{z_{\alpha/2}^2}{w^2}=\frac{(1.645)^2}{0.04^2}=1691.3 \qquad \Longrightarrow \qquad n\ge\frac{4z_{\alpha/2}^2\hat{p}\hat{q}}{w^2}.$$

Thus about $n=1692$ observations would imply a confidence interval width of 0.04 or less.