**1.** The dividend yields for all companies whose shares were traded on the New York Stock Exchange in 1990 obeyed a normal distribution with standard deviation 2.6%. A random sample of 16 of these companies was taken in order to estimate the population mean dividend yield.

(a) (6 pts) What is the probability that the sample mean was below the population mean by more than 1.1%?

Population has $N(\mu, (2.6\%)^2)$ distribution.

Sample mean $\bar{X} \sim N\left(\mu, \frac{(2.6\%)^2}{16}\right)$

$$P(\bar{X} < \mu - 1.1) = P(\bar{X} - \mu < -1.1) = P\left(\frac{\bar{X}-\mu}{2.6/4} < \frac{-1.1}{2.6/4}\right)$$

$$= P(Z < -1.69) \quad \text{where } Z = \frac{\bar{X}-\mu}{2.6/4} \sim N(0,1)$$

$= .0455$ using the normal table.

(b) (3 pts) Suppose that a second (independent) random sample of 40 companies was taken. Without calculation, state whether the probability in (a) would be higher, lower, or the same for the second sample.

As the sample size increases, the variance of $\bar{X}$ (sample mean) decreases and $P(\bar{X} - \mu < -1.1)$ also decreases because $\bar{X} - \mu \sim N(0, \frac{\sigma^2}{n})$ where $\frac{\sigma^2}{n}$ is the variance of $\bar{X}$.

So the probability in (a) would be lower for the second sample.

(c) (6 pts) Suppose you did not know the population standard deviation, instead you knew only the standard deviation of the dividend yield for the 16 randomly selected companies as 2.6%. Find the probability that the sample mean was below the population mean by more than 1.1%.

Since the population is normal with an unknown standard deviation, $\frac{\bar{X}-\mu}{s/\sqrt{n}}$ will have a t distribution with $n-1$ d.f. and we will use it to find the probability.

$$P(\bar{X} - \mu < -1.1) = P\left(\frac{\bar{X}-\mu}{2.6/\sqrt{16}} < \frac{-1.1}{2.6/4}\right)$$

$$= P(T < -1.69) \quad \text{where } T = \frac{\bar{X}-\mu}{2.6/\sqrt{16}} \sim t \text{ with } 15 \text{ d.f.}$$

$P(T < -1.69)$ ~~is~~ in between .05 and .1

**2.** Let $X$ and $Y$ be two random variables with $V(X) = 5, V(Y) = 20, Cov(X,Y) = -6$.

**(a)** (6 pts) Find $V(X+Y)$.

$$
\begin{aligned}
V(X+Y) &= V(X) + V(Y) + 2Cov(X,Y) \\
&= 5 + 20 + 2\cdot(-6) \\
&= 13
\end{aligned}
$$

**(b)** (6 pts) Find $V(X-2Y)$.

$$
\begin{aligned}
V(X-2Y) &= V(X) + 4V(Y) - 2\cdot 2\, Cov(X,Y) \\
&= 5 + 4\cdot 20 - 4\cdot(-6) \\
&= 109
\end{aligned}
$$

**(c)** (6 pts) Find $Cov(X-2Y, X+Y)$.

$$
\begin{aligned}
&Cov(X-2Y, X+Y) \\
&= Cov(X,X) - 2Cov(Y,X) + Cov(X,Y) - 2Cov(Y,Y) \\
&= V(X) - Cov(X,Y) - 2V(Y) \\
&= 5 - (-6) - 2\cdot 20 \\
&= -29
\end{aligned}
$$

**(d)** (3 pts) Find the correlation coefficient between $X - 2Y$ and $X + Y$.

correlation coefficient between $X-2Y$ and $X+Y$ is $\dfrac{Cov(X-2Y, X+Y)}{\sqrt{V(X-2Y)\,V(X+Y)}} = \dfrac{-29}{\sqrt{109\cdot 13}} \approx -0.77$

**(e)** (3 pts) Find the correlation coefficient between $2X - 4Y - 4$ and $4X + 4Y + 10$.

Correlation coefficient between $2X-4Y-4$ and $4X+4Y+10$ is same as the correlation coefficient between $X-2Y$ and $X+Y$ because $2X-4Y-4 = 2(X-2Y) - 4$ and $4X+4Y+10 = 4(X+Y)+10$.

The answer is $-0.77$

**3.** The joint probability density function of two random variables $X$ and $Y$ is given by

$$f(x,y) = \begin{cases} \frac{1}{10}(x^2+y^2) & \text{if } 0 \le x \le 1,\ 0 \le y \le 3 \\ 0 & \text{otherwise.} \end{cases}$$

(a) (9 pts) Find the probability that $X \le .5$ and $Y \le 1$.

$$P(X \le .5, Y \le 1) = \int_0^{.5}\int_0^{1} \frac{1}{10}(x^2+y^2)\,dy\,dx = \frac{1}{10}\int_0^{.5}\left[x^2y + \frac{y^3}{3}\right]_0^1 dx$$

$$= \frac{1}{10}\int_0^{.5}\left(x^2+\frac{1}{3}\right)dx$$

$$= \frac{1}{10}\left[\frac{x^3}{3}+\frac{x}{3}\right]_0^{.5}$$

$$= \frac{1}{10}\left(\frac{.125}{3}+\frac{.5}{3}\right)$$

$$= 0.0208$$

(b) (3 pts) Evaluate the joint cumulative distribution function of $X$ and $Y$ at (.5, 1).

Using the definition of the joint c.d.f. we have

$F(.5, 1) = P(X \le .5, Y \le 1) = .0208$ from (a). Here $F(x,y)$ is the joint c.d.f. of $(X,Y)$

(c) (9 pts) Find the expected value of $X^2 - Y^2$.

$$E(X^2-Y^2) = \int_0^1\int_0^3 (x^2-y^2)\frac{1}{10}(x^2+y^2)\,dy\,dx$$

$$= \frac{1}{10}\int_0^1\int_0^3 (x^4-y^4)\,dy\,dx = \frac{1}{10}\int_0^1\left[x^4y - \frac{y^5}{5}\right]_0^3 dx$$

$$= \frac{1}{10}\int_0^1\left(3x^4 - \frac{3^5}{5}\right)dx$$

$$= \frac{1}{10}\left[3\frac{x^5}{5} - \frac{3^5}{5}\cdot x\right]_0^1$$

$$= \frac{1}{10}\left(\frac{3}{5}-\frac{3^5}{5}\right)$$

$$= \frac{1}{10}\cdot\frac{3}{5}(1-81)$$

$$= -\frac{24}{5}$$

$$= -4.8$$

(d) (12 pts) Find the conditional probability that $X \leq .5$ given that $Y = 2$.

First we have to find the conditional density of $X$ given $Y=2$ and then integrate over the interval $[0, .5]$

The conditional density of $X \mid Y=2$ is $\frac{f(x,y)}{f_2(y)}$ evaluated at $y=2$.

Here $f_2(y)$ is the marginal density of $Y$.

$$f_2(y) = \int_0^1 \frac{1}{10}(x^2+y^2)\,dx = \frac{1}{10}\left[\frac{x^3}{3} + xy^2\right]_0^1 = \frac{1}{10}\left(\frac{1}{3} + y^2\right)$$

The conditional density of $X \mid Y=y$ is $\frac{\frac{1}{10}(x^2+y^2)}{\frac{1}{10}\left(\frac{1}{3}+y^2\right)} = \frac{x^2+y^2}{\frac{1}{3}+y^2}$

The conditional density of $X \mid Y=2$ is $\frac{x^2+4}{\frac{1}{3}+4} = \frac{3}{13}(x^2+4)$

$$P(X \leq .5 \mid Y=2) = \int_0^{.5} \frac{3}{13}(x^2+4)\,dx$$

$$= \frac{3}{13}\left[\frac{x^3}{3} + 4x\right]_0^{.5}$$

$$= \frac{3}{13}\left(\frac{.125}{3} + 4 \cdot .5\right)$$

$$= \frac{3}{13} \cdot \frac{6.125}{3}$$

$$= \frac{6.125}{13}$$

$$\approx .471$$

**4.** Let $X_1, X_2, X_3, X_4$ be a random sample from a large population with mean $\mu$ and variance $\sigma^2$. Two estimators $\hat{\mu}_1$ and $\hat{\mu}_2$ for the parameter $\mu$ are defined as

$$\hat{\mu}_1 = \frac{X_1 + X_2 + X_3 + X_4}{4}$$
$$\hat{\mu}_2 = \frac{X_2 + X_3 + X_4}{3}$$

(a) (5 pts) Find the bias for $\hat{\mu}_1$ and $\hat{\mu}_2$.

Note that $\hat{\mu}_1$ is the sample mean based on the sample $X_1, X_2, X_3, X_4$ and $\hat{\mu}_2$ is the sample mean based on the sample $X_2, X_3, X_4$

So, $E(\hat{\mu}_1) = \mu$ and $E(\hat{\mu}_2) = \mu$

Therefore, $\text{Bias}(\hat{\mu}_1) = E(\hat{\mu}_1) - \mu = \mu - \mu = 0$

and $\text{Bias}(\hat{\mu}_2) = E(\hat{\mu}_2) - \mu = \mu - \mu = 0$

(b) (9 pts) Which one of $\hat{\mu}_1$ and $\hat{\mu}_2$ is a better estimator with respect to the variance criterion?

$V(\hat{\mu}_1) = \frac{\sigma^2}{4}$ and $V(\hat{\mu}_2) = \frac{\sigma^2}{3}$ since $\hat{\mu}_1$ and $\hat{\mu}_2$ are sample means based on samples of size 4 and 3, respectively.

The estimator with a smaller variance is better. So $\hat{\mu}_1$ is a better estimator with respect to the variance criterion.

**5.** Seventy-five percent of a school's law class passes the state bar examination on the first attempt.

(a) (7 pts) If a randomly selected group of 250 of this school's law graduates take the state bar examination, what is the probability that 80% or more of them will pass the examination on the first attempt?

Population proportion $p = .75$

Let $\hat{p}$ denote the sample proportion.

Sample size $n = 250$

By CLT, $\hat{p} \sim N(p, \frac{pq}{n}) = N(.75, .00075)$

$$P(\hat{p} \geq .8) = P\left(\frac{\hat{p} - .75}{\sqrt{.00075}} \geq \frac{.8 - .75}{\sqrt{.00075}}\right)$$

$= P(Z \geq 1.83)$ where $Z = \frac{\hat{p} - .75}{\sqrt{.00075}} \sim N(0,1)$

$= .0336$

(b) (7 pts) If a randomly selected group of 120 of this school's law graduates take the state bar examination, what is the probability that exactly 100 of them will pass the examination on the first attempt?

Let $Y$ be the number of students who pass the exam.

Then $Y \sim \text{Binomial}(120, .75)$

By CLT $Y \sim N(np, npq)$ where $n = 120, p = .75, q = 1-p$

$= N(90, 22.5)$

By continuity correction,

$$P(Y = 100) = P(99.5 < Y < 100.5) = P\left(\frac{99.5 - 90}{\sqrt{22.5}} < \frac{Y - 90}{\sqrt{22.5}} < \frac{100.5 - 90}{\sqrt{22.5}}\right)$$

$= P(2 < Z < 2.21)$ where $Z = \frac{Y - 90}{\sqrt{22.5}} \sim N(0,1)$

$= 0.0228 - 0.0136$

$= 0.0092$