Problem 9.1 (**Video 8.1, 8.2, 9.1, 9.2**, **Lecture Problem**)

Let $X_1, \ldots, X_n$ be i.i.d. random variables with PMF

$$P_X(x) = \begin{cases} 1/2 & x = 2 \\ 1/2 & x = 4 \\ 0 & \text{otherwise.} \end{cases}$$

Let $S_{300} = X_1 + \cdots + X_{300}$.

(a) Determine the mean of and variance of S_{300}.

(b) Use the Central Limit Theorem approximation to estimate the probability $\mathbb{P}\big[\big|S_{300} - E[S_{300}]\big| \geq 40\big]$. You can leave your answer in terms of the standard normal CDF $\Phi(z)$.

(c) Suppose you did not know the PMF $P_X(x)$, but you knew that the standard deviation of the random variables X_k was one. You measure $X_1, \ldots, X_{300}$ and compute the sample mean $M_{300} = \frac{1}{300}\sum_{k=1}^{300} X_k$. Find a symmetric confidence interval for the true mean around the observed value M_{300} with confidence level 0.95.

Use the following assumptions: $Q(1.28) = 1 - \Phi(1.28) = 0.1$; $Q(1.645) = 1 - \Phi(1.645) = 0.05$; $Q(1.96) = 1 - \Phi(1.96) = 0.025$.

(d) Suppose you also did not know the standard deviation, but you were able to compute the sample variance of the 300 samples X_k as $V_{300} = 1.21$. Find a symmetric confidence interval for the true mean around the observed value M_{300} with confidence level 0.95.

Use the following assumptions: if W has a t-distribution with 299 degrees of freedom, its CDF satisfies: $F_W(-1.9679) = 0.025$; $F_W(-1.65) = 0.05$; $F_W(-1.2844) = 0.05$.

Problem 9.2 (Video 8.1, 8.2, 9.1, 9.2)

You are working with an immortal cell line that is used in labs across the world and known to have a mean radius of $5.02\mu m$ with a standard deviation of $0.50\mu m$. You collect 100 samples, and assume that these are well-modeled as i.i.d. Gaussian random variables.

You may find one or more of the following values useful: $Q(1.28) = 1 - \Phi(1.28) = 0.1$; $Q(1.645) = 1 - \Phi(1.645) = 0.05$; $Q(1.96) = 1 - \Phi(1.96) = 0.025$.

(a) What is the variance of the sample mean $\mathsf{Var}[M_{100}]$?

(b) If your sample mean is $5.10\mu m$ is your sample significantly different from the baseline model at a significance level of 0.05? Justify your approach and support your answer numerically.

(c) Say you do not know the true mean (but you still know the true standard deviation is $0.50\mu m$). You observe a sample mean of $5.10\mu m$. Construct a confidence interval centered around the sample mean with confidence level 0.9 for the true mean.

a) $\mathrm{Var}(\bar{x}) = \frac{1}{10^4} \sum_{i=1}^{100} \frac{x_i}{100} = \frac{1}{10^4} (100 \cdot 0.5^2)\ \mu m^2 = 0.25 \cdot 10^{-2}\ \mu m^2$

b) $\bar{x} = 5.1 \mu m$

$\mu = 5.02 \mu m$

$\frac{\bar{x} - \mu}{sd/\sqrt{n}} = \frac{5.1 - 5.02}{0.5} = \frac{0.08 \cdot 10}{0.5} = 1.6$

$P(z > 1.6) = 0.0548 > 0.05$

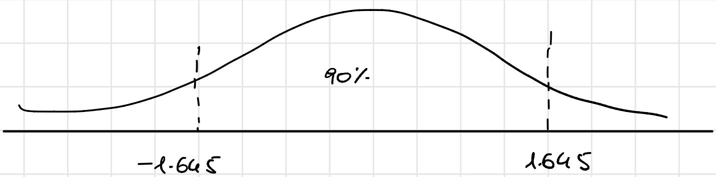
$\downarrow$

Meaning the sample mean is statistically significant

c) $\bar{x} = 5.1 \mu m$

$\sigma = 0.5 \mu m$

$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$



$P\left(-1.645 \le \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \le 1.645\right) = 0.9$

$P\left(\bar{x} - 1.645 \frac{\sigma}{\sqrt{n}} \le \mu < \bar{x} + 1.645 \frac{\sigma}{\sqrt{n}}\right) = 0.9$

$P(5.0178 \le \mu \le 5.1823) = 0.9$

90% is $(5.0178, 5.1823)$

Problem 9.3 (Video 9.1, 9.2, 9.3, 9.4 Lecture Problem)

You measure the sulfate concentration in the local water reservoir over 9 consecutive days and obtain values $X_1, \ldots, X_9$, which are assumed to be i.i.d. Gaussian. (The units are mg/L and omitted below.) The sample mean is $M_9 = 6.1$ and the sample variance is $V_9 = 0.36$.

Let W have a t-distribution with 8 degrees-of-freedom. You can assume the following values are available:

- $F_W(-1.4) = \Phi(-1.3) = Q(1.3) = 0.1, \ F_W(1.4) = \Phi(1.3) = 0.9$
- $F_W(-1.9) = \Phi(-1.6) = Q(1.6) = 0.05, \ F_W(1.9) = \Phi(1.6) = 0.95$
- $F_W(-2.3) = \Phi(-2.0) = Q(2.0) = 0.025, \ F_W(2.3) = \Phi(2.0) = 0.975$

(a) Construct a confidence interval for the mean with confidence level 0.9.

(b) Is your sample significantly different from the baseline concentration $\mu = 5.4$ at a significance level of 0.05? Justify your approach and support your answer numerically.

(c) Say you also go out on the 10^{th} day and collect measurement $X_{10} = 5$. What is the new sample mean M_{10}?

a) $\left(M_9 - t_{\alpha/2, n-1}\sqrt{\frac{V_9}{n}}, \ M_9 + t_{\alpha/2, n-1}\sqrt{\frac{V_9}{n}}\right)$

$\left(6.1 - 1.9\sqrt{\frac{0.36}{9}}, \ 6.1 + 1.9\sqrt{\frac{0.36}{9}}\right)$

$(5.72, 6.48)$

b) $\left(6.1 - 2.3\sqrt{\frac{0.36}{9}}, \ 6.1 + 2.3\sqrt{\frac{0.36}{9}}\right)$

$(5.64, 6.56)$

c) $M_{10} = \frac{9 \cdot M_9 + 5}{10} = \frac{9 \cdot 6.1 + 5}{10} = 5.99$

Problem 9.4 (**Video 7.1, 7.2**, **Quick Calculations**) For each of the scenarios below, determine the requested quantities. (You should be able to do this without any long calculations or integration.)

(a) Assume that X and Y are jointly Gaussian with $\mathbb{E}[X] = 1$, $\mathbb{E}[Y] = 2$, $\mathsf{Var}[X] = 1$, $\mathsf{Var}[Y] = 4$, $\rho_{X,Y} = -\frac{1}{2}$. Determine MMSE estimator of X given $Y = y$ and the corresponding mean-squard error.

(b) Assume that X is Uniform$(1, 3)$, V is Gaussian$(1, 1)$, X and V are independent, and $Y = X + V$. Determine the LLSE estimator of X given $Y = y$ and the corresponding mean-squared error.

a) $$X_{MMSE}(y) = E(X \mid Y=y) = \mu_X + \rho_{XY} + \frac{\sigma_X}{\sigma_Y}(y - \mu_Y)$$

$$= 1 + \left(-\tfrac{1}{2}\right)\left(\tfrac{1}{2}\right)(y-2) = \boxed{\frac{3}{2} - \frac{1}{4}y}$$

$$MSE = (1 - \rho_{XY}^2)\,\sigma_X^2 = \left(1 - \tfrac{1}{4}\right)1 = \boxed{\frac{3}{4}}$$

b) $$X_{LLSE}(y) = E(X) + \frac{\mathrm{cov}\,X,Y}{\mathrm{var}\,Y}(y - E(Y))$$

$\mu_Y = 1 + 2 = 3$

$\sigma_Y = 3 + 1 = 4/3$

$\sigma_X = \frac{4}{12} = \frac{1}{3}$

$$= 2 + \frac{\mathrm{cov}\,X,Y}{4}(y - 3)$$

$$= 2 + \frac{1/3}{4/3}(y-3)$$

$$= 2 + \frac{1}{4}(y-3) = 2 + \frac{y}{4} - \frac{3}{4} = \boxed{\frac{y}{4} - \frac{5}{4}}$$

$$MSE_{LLSE} = \mathrm{Var}\,X - \frac{\mathrm{Cov}\,XY^2}{\mathrm{var}\,Y} = \frac{1}{3} - \frac{1/9}{4/3} = \boxed{1/4}$$