Gov 2001: Midterm Exam

Spring 2025

March , 2025

Midterm Instructions:

- This midterm exam is due on **March 14, 11:59 pm** Eastern time. Please upload a PDF of your solutions to **Gradescope**.
- We will accept hand-written solutions but we strongly advise graduate students to typeset your answers in LaTeX.
- This is a semi-closed book test. You are **NOT** allowed to: search internet / AI for solutions or communicate amongst each other.
- You are allowed to utilize class materials (slides, section slides, pset solutions).

1 Variance and Covariance (25pt)

- Write down the definition of Covariance and interpret with no more than 2 sentences. (5pt)
- Give a counter example when X and Y are dependent but with a covariance of zero. (5pt)

See slides/textbook for the definition and interpretation. Counter example: $Y = X^2$ where $X \sim \mathcal{N}(0,1)$ or Unif$(-1,1)$, i.e., a distribution symmetric to zero, which makes odd moments zero. They are clearly dependent since knowing X gives you all the information about Y, but $\text{cov}(X,Y) = E[X^3] - E[X]E[X^2] = 0 - 0 = 0$.

1.1 Prove the following:

(5pt each)

- $\text{cov}(X,X) = V(X)$
- $\text{cov}(X+Y, Z+W) = \text{cov}(X,Z) + \text{cov}(X,W) + \text{cov}(Y,Z) + \text{cov}(Y,W)$
- $V(X+Y) = V(X) + V(Y) + 2\text{cov}(X,Y)$

- $\text{cov}(X, X) = E[X^2] - E[X]E[X] = V(X)$
- We should use the definition of covariance and linearity of expectation:

$$\begin{aligned}\text{cov}(X+Y, Z+W) &= E[(X+Y)(Z+W)] - E[X+Y]E[Z+W] \\ &= E[XZ] + E[XW] + E[YZ] + E[YW] \\ &- E[X]E[Z] - E[X]E[W] - E[Y]E[Z] - E[Y]E[W] \\ &= \text{cov}(X, Z) + \text{cov}(X, W) + \text{cov}(Y, Z) + \text{cov}(Y, W)\end{aligned}$$

- Use the above two conclusions:

$$\begin{aligned}V(X+Y) &= \text{cov}(X+Y, X+Y) \\ &= \text{cov}(X, X) + 2\text{cov}(X, Y) + \text{cov}(Y, Y) \\ &= V(X) + V(Y) + 2\text{cov}(X, Y)\end{aligned}$$

2 Correlation (30pt)

- Write down the definition of Correlation between two r.v.s X_1 and X_2 and show why it's always in between -1 and 1. (10pt)

  Hints: you will need to utilize Cauchy-Schwarz inequality: For any real-valued random variables X and Y),

$$|E[AB]| \leq \sqrt{E[A^2]E[B^2]}.$$

  The correlation between two random variables X_1 and X_2 is given by the Pearson correlation coefficient:

$$\rho(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}}.$$

  where:

  - $\text{Cov}(X_1, X_2) = E[(X_1 - E[X_1])(X_2 - E[X_2])]$ is the covariance between X_1 and X_2,
  - $\sigma_{X_1} = \sqrt{\text{Var}(X_1)}$ and $\sigma_{X_2} = \sqrt{\text{Var}(X_2)}$ are the standard deviations of X_1 and X_2, respectively.

  We start with the **Cauchy-Schwarz inequality**, which states that for any two random variables X_1 and X_2:

$$|E[(X_1 - E[X_1])(X_2 - E[X_2])]| \leq \sqrt{E[(X_1 - E[X_1])^2]E[(X_2 - E[X_2])^2]}.$$

  Since the left-hand side is the absolute value of $\text{Cov}(X_1, X_2)$ and the right-hand side is $\sigma_{X_1}\sigma_{X_2}$, we obtain:

$$|\text{Cov}(X_1, X_2)| \leq \sigma_{X_1}\sigma_{X_2}.$$

Dividing both sides by $\sigma_{X_1}\sigma_{X_2}$ (assuming both standard deviations are nonzero):

$$\left|\frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}}\right| \leq 1.$$

Since the correlation coefficient is defined as:

$$\rho(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}},$$

we conclude:

$$-1 \leq \rho(X_1, X_2) \leq 1.$$

- Let's say you are studying the returns of two stocks, X_1 and X_2. Suppose the daily returns of these stocks are normally distributed with means μ_1 and μ_2, variances σ_1^2 and σ_2^2, and correlation ρ, meaning:

$$\text{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2.$$

1. (10pt) Derive the expected return and variance of a portfolio consisting of an equal-weighted combination of these two stocks, i.e.,

$$S = \frac{X_1 + X_2}{2}.$$

The expected return of the portfolio is:

$$E[S] = E\left[\frac{X_1 + X_2}{2}\right] = \frac{E[X_1] + E[X_2]}{2} = \frac{\mu_1 + \mu_2}{2}.$$

The variance of the portfolio is:

$$\text{Var}(S) = \text{Var}\left(\frac{X_1 + X_2}{2}\right).$$

Using the variance properties:

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y),$$

we substitute $a = b = \frac{1}{2}$:

$$\text{Var}(S) = \left(\frac{1}{2}\right)^2 \sigma_1^2 + \left(\frac{1}{2}\right)^2 \sigma_2^2 + 2 \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \rho\sigma_1\sigma_2.$$

$$= \frac{\sigma_1^2 + 2\rho\sigma_1\sigma_2 + \sigma_2^2}{4}.$$

2. Let's try to standardize the stock returns using:

$$Z_1 = \frac{X_1 - \mu_1}{\sigma_1}, \quad Z_2 = \frac{X_2 - \mu_2}{\sigma_2}.$$

(a) Compute the correlation coefficient between Z_1 and Z_2 (5pt).
After standardization, we have $\sigma_{Z_1} = \sigma_{Z_2} = 1$, and $\mu_{Z_1} = \mu_{Z_2} = 0$. Therefore:

$$\begin{aligned} \text{Corr}(Z_1, Z_2) &= \frac{E[Z_1 Z_2] - \mu_{Z_1}\mu_{Z_2}}{\sigma_{Z_1}\sigma_{Z_2}} \\ &= E[Z_1 Z_2] \\ &= \frac{1}{\sigma_1\sigma_2} E[(X_1 - \mu_1)(X_2 - \mu_2)] \\ &= \frac{\text{Cov}(X_1, X_2)}{\sigma_1\sigma_2} \\ &= \frac{\rho\sigma_1\sigma_2}{\sigma_1\sigma_2} = \rho \end{aligned}$$

Or you can directly use the property: since standardization does not affect the correlation (see section 4 notes), we have:

$$\text{Corr}(Z_1, Z_2) = \rho.$$

(b) Explain the significance of this correlation in the stock returns context with no more than 2 sentences. (5pt)

- Correlation between standardized returns helps in risk assessment (low or zero correlation helps reduce potential fluctuations in returns).
- Even when returns are measured on different scales, their standardized correlation remains unchanged.
- Portfolio diversification benefits depend on correlation, not just individual variances.

3 "Tea" Testing and Forecast (30pt)

In this problem, we're going to explore a real-world example of Fisher's "lady tasting tea" experiment from lecture: election forecasters – who have, for better or worse, become a big part of politics in the United States and elsewhere.

- Suppose that Bob has correctly predicted six of the last eight election outcomes. What is the probability that someone randomly flipping a coin each of the same elections would have experienced *at least* the same success as Bob? Compute your answer analytically (i.e. not by simulation).(10pt)

  Bob has correctly predicted 6 out of the last 8 election outcomes. If a person were randomly flipping a fair coin for each election, the probability of guessing correctly follows a binomial distribution:

$$X \sim \text{Binomial}(n = 8, p = 0.5).$$

We need to compute:

$$P(X \geq 6) = P(X = 6) + P(X = 7) + P(X = 8).$$

Using the binomial probability mass function:

$$P(X = k) = \binom{8}{k}(0.5)^8.$$

Computing each term:

$$P(X = 6) = \binom{8}{6}(0.5)^8 = \frac{8!}{6!(8-6)!}(0.5)^8 = \frac{28}{256} = 0.1094.$$

$$P(X = 7) = \binom{8}{7}(0.5)^8 = \frac{8!}{7!(8-7)!}(0.5)^8 = \frac{8}{256} = 0.0313.$$

$$P(X = 8) = \binom{8}{8}(0.5)^8 = \frac{1}{256} = 0.0039.$$

Summing these probabilities:

$$P(X \geq 6) = 0.1094 + 0.0313 + 0.0039 = 0.1445.$$

Thus, the probability that a random guesser achieves Bob's success is approximately **0.1445**.

- Forecasting has become so popular that riffraff are flooding the market. These "uniform amateurs" predict the vote share for each state in the U.S. presidential election by drawing a uniform random variable between 0 and 1, independently across states. You are deciding whether or not to hire a forecaster, Nate, to forecast each of the 50 state election winners in the 2024 presidential general election based on the performance of his 2020 election forecast, but you are worried that Nate might be one of these amateurs. When you ask him to justify his 2020 forecasts, he says "my highest predicted [Democratic] vote share was 0.8 which is very unlikely if I were a uniform amateur." Let's evaluate his claim.

  Suppose Nate is a uniform amateur and let X be the maximum of the 51 uniform vote share draws (include D.C.). Derive the CDF and PDF of X. Use these to calculate the probability of Nate's highest Democratic vote share being 0.8 or less if he were a uniform amateur. (10pt)

  If Nate is a uniform amateur, he assigns a vote share $U_i \sim \text{Uniform}(0, 1)$ independently across the 51 regions (50 states + D.C.). Define:

$$X = \max(U_1, U_2, \dots, U_{51}).$$

The cumulative distribution function (CDF) of X is:

$$P(X \leq x) = P(U_1 \leq x, U_2 \leq x, \dots, U_{51} \leq x).$$

Since the U_i are independent:

$$P(X \leq x) = P(U_1 \leq x)P(U_2 \leq x)\dots P(U_{51} \leq x).$$

Since $P(U_i \leq x) = x$ for $U_i \sim \text{Uniform}(0,1)$:

$$F_X(x) = x^{51}, \quad \text{for } 0 \leq x \leq 1.$$

The probability density function (PDF) is obtained by differentiating:

$$f_X(x) = \frac{d}{dx}x^{51} = 51x^{50}, \quad \text{for } 0 \leq x \leq 1.$$

We now compute:

$$P(X \leq 0.8) = F_X(0.8) = (0.8)^{51}.$$

Using numerical computation:

$$(0.8)^{51} \approx 0.00017.$$

Thus, the probability that an amateur's highest predicted vote share is 0.8 or less is **0.00017**, which is very small, suggesting that Nate's claim is reasonable.

- To be on the lookout for more uniform amateurs, it's helpful to know what highest vote share we should expect. To that end, calculate $E[X]$.(10pt)

  To compute $E[X]$, we use:

$$E[X] = \int_0^1 x f_X(x) dx.$$

Substituting $f_X(x) = 51x^{50}$:

$$E[X] = \int_0^1 x(51x^{50}) dx.$$

$$= 51 \int_0^1 x^{51} dx.$$

Evaluating the integral:

$$\int x^{51} dx = \frac{x^{52}}{52}\Big|_0^1 = \frac{1}{52}.$$

Thus,

$$E[X] = 51 \times \frac{1}{52} = \frac{51}{52} \approx 0.9808.$$

The expected maximum vote share for a uniform amateur is **0.9808**.

4 Normal Distribution (15pt + bonus 10pt)

Let a standard normal r.v to be $Z \sim \mathcal{N}(0,1)$.

- Express the random variable $Y \sim \mathcal{N}(1,2)$ as a simple function in terms of Z. Make sure to check that your Y has the correct mean and variance. (5pt)

  A general normal random variable $Y \sim \mathcal{N}(\mu, \sigma^2)$ can be expressed in terms of a standard normal variable $Z \sim \mathcal{N}(0,1)$ as:

  $$Y = \mu + \sigma Z.$$

  For $Y \sim \mathcal{N}(1,2)$, we identify $\mu = 1$ and $\sigma^2 = 2$, meaning $\sigma = \sqrt{2}$. Thus, we express Y as:

  $$Y = 1 + \sqrt{2}Z.$$

  Verification: To confirm that $Y \sim \mathcal{N}(1,2)$:

  $$E[Y] = E[1 + \sqrt{2}Z] = 1 + \sqrt{2}E[Z] = 1 + 0 = 1.$$

  $$\mathrm{Var}(Y) = \mathrm{Var}(1 + \sqrt{2}Z) = \mathrm{Var}(\sqrt{2}Z) = (\sqrt{2})^2\mathrm{Var}(Z) = 2 \times 1 = 2.$$

  Thus, the mean and variance match the given distribution, confirming the correctness.

- Express the probability of $|Y| \leq 1$ as a function of Φ, the CDF of the standard Normal distribution.(10pt)

  We need to express:

  $$P(|Y| \leq 1) = P(-1 \leq Y \leq 1)$$

  Using the transformation $Y = 1 + \sqrt{2}Z$, we rewrite the probability:

$$P(-1 \le 1 + \sqrt{2}Z \le 1).$$

Subtracting 1 from all sides:

$$P(-2 \le \sqrt{2}Z \le 0).$$

Dividing by $\sqrt{2}$:

$$P\left(\frac{-2}{\sqrt{2}} \le Z \le \frac{0}{\sqrt{2}}\right).$$

Since $\frac{-2}{\sqrt{2}} = -\sqrt{2}$ and $\frac{0}{\sqrt{2}} = 0$, this simplifies to:

$$P(-\sqrt{2} \le Z \le 0).$$

Using the CDF $\Phi(x)$, we express this probability as:

$$\Phi(0) - \Phi(-\sqrt{2}).$$

Since $\Phi(0) = 0.5$, we get:

$$P(|Y| \le 1) = 0.5 - \Phi(-\sqrt{2}).$$

Using the symmetry property $\Phi(-x) = 1 - \Phi(x)$, we rewrite it as:

$$P(|Y| \le 1) = \Phi(\sqrt{2}) - 0.5.$$

Bonus Q: Prove that: If $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and $X_1 \perp\!\!\!\perp X_2$,

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

(optional, 10pt)

The moment-generating function (MGF) of a random variable X, denoted $M_X(t)$, is defined as:

$$M_X(t) = E[e^{tX}].$$

For a normal random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, let $X = \mu + \sigma Z$, where $Z \sim \mathcal{N}(0,1)$. The MGF of Z is, by LOTUS,

$$E[e^{tZ}] = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x-t)^2 + \frac{t^2}{2}\right) dx = \exp\left(\frac{t^2}{2}\right)$$

Since we can consider the following as the integral of a standard normal PDF on its support:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-t)^2\right) d(x-t) = 1$$

Therefore, the MGF of X is

$$M_X(t) = E[e^{t(\mu+\sigma Z)}] = e^{t\mu} E[e^{(t\sigma)Z}] = e^{t\mu} \exp\left(\frac{1}{2}\sigma^2 t^2\right) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right).$$

Applying this to $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, their MGFs are:

$$M_{X_1}(t) = \exp\left(\mu_1 t + \frac{1}{2}\sigma_1^2 t^2\right),$$

$$M_{X_2}(t) = \exp\left(\mu_2 t + \frac{1}{2}\sigma_2^2 t^2\right).$$

Then, let's take a look at the MGF of $S = X_1 + X_2$

Since X_1 and X_2 are independent, the MGF of their sum satisfies:

$$M_S(t) = M_{X_1+X_2}(t) = M_{X_1}(t) M_{X_2}(t).$$

Substituting the MGFs:

$$M_S(t) = \left[\exp\left(\mu_1 t + \frac{1}{2}\sigma_1^2 t^2\right)\right] \times \left[\exp\left(\mu_2 t + \frac{1}{2}\sigma_2^2 t^2\right)\right].$$

Using the property $e^a e^b = e^{a+b}$, we simplify:

$$M_S(t) = \exp\left((\mu_1 + \mu_2)t + \frac{1}{2}(\sigma_1^2 + \sigma_2^2)t^2\right).$$

Comparing with the known MGF of a normal distribution:

$$M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right),$$

we see that $M_S(t)$ matches the form of an MGF for a normal distribution with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. Thus,

$$X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$