STA 130A: Mathematical Statistics: Brief Course

Practice midterm exam 2 — Set A solutions

Instructor: Dogyoon Song

Problem 1. Conceptual checks

(a) An affine transformation. The map $Z = 3 - 2U$ sends $(0, 1)$ to $(1, 3)$; its inverse has absolute derivative $1/2$. Thus

$$f_Z(z) = \begin{cases} \frac{1}{2}, & 1 < z < 3, \\ 0, & \text{otherwise.} \end{cases}$$

(b) Zero correlation versus independence. No: zero correlation means zero covariance, which does not rule out nonlinear dependence and does not imply factorization of the joint distribution.

(c) Variance of a linear combination.

$$\mathrm{Var}(2X - Y) = 4\mathrm{Var}(X) + \mathrm{Var}(Y) - 4\mathrm{Cov}(X, Y) = 4 + 4 - 4 = 4.$$

(d) Variance within one group. No: for example, two equally centered groups with variances $4, 0$ and probabilities $1/4, 3/4$ have overall variance 1, below the first group's variance 4. Total variance controls the *average* conditional variance, not every group's variance.

(e) A second moment from an MGF. For $t < 1/2$,

$$M_X'(t) = 2(1 - 2t)^{-2}, \qquad M_X''(t) = 8(1 - 2t)^{-3}.$$

Since the MGF is finite near zero, $\mathbb{E}[X^2] = M_X''(0) = 8$.

(f) Markov applied to a square. Since $X^2 \geq 0$,

$$P(|X| \geq 6) = P(X^2 \geq 36) \leq \frac{\mathbb{E}[X^2]}{36} = \frac{9}{36} = \frac{1}{4}.$$

Problem 2. A joint density on a triangle

The joint density is 1 on $0 < x < 1$, $0 < y < 2(1 - x)$, and zero elsewhere.

(a) Marginal density of X. For $0 < x < 1$,

$$f_X(x) = \int_0^{2(1-x)} 1 \, dy = 2(1 - x).$$

Therefore

$$f_X(x) = \begin{cases} 2(1 - x), & 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

(b) Conditional probability. Using the conditional-density formula at $x = 1/2$,

$$f_{Y|X}(y \mid 1/2) = \frac{f_{X,Y}(1/2, y)}{f_X(1/2)} = \begin{cases} 1, & 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Consequently,

$$P(Y > 1/2 \mid X = 1/2) = \int_{1/2}^{1} 1 \, dy = \frac{1}{2}.$$

(c) A probability over a region. Inside the joint support, $X + Y/2 < 1/2$ is equivalent to $0 < x < 1/2$, $0 < y < 1 - 2x$. This smaller triangle lies inside the original support. Hence

$$P(X + Y/2 < 1/2) = \int_0^{1/2} \int_0^{1-2x} 1 \, dy \, dx = \int_0^{1/2} (1 - 2x) \, dx = \frac{1}{4}.$$

Problem 3. MGFs and correlation of linear combinations

(a) MGF of R. From the PMF,

$$M_R(t) = \mathbb{E}[e^{tR}] = \frac{1}{2}e^{-t} + \frac{1}{4} + \frac{1}{4}e^{2t}, \qquad t \in \mathbb{R}.$$

The finite support makes this finite for every real t.

(b) MGF of $S = 3 + R_1 - 2R_2$. Independence gives

$$\begin{aligned} M_S(t) &= e^{3t}\mathbb{E}[e^{tR_1}e^{-2tR_2}] \\ &= e^{3t}\mathbb{E}[e^{tR_1}]\mathbb{E}[e^{-2tR_2}] = e^{3t}M_R(t)M_R(-2t). \end{aligned}$$

Substituting the explicit MGF,

$$M_S(t) = e^{3t}\left(\frac{1}{2}e^{-t} + \frac{1}{4} + \frac{1}{4}e^{2t}\right)\left(\frac{1}{2}e^{2t} + \frac{1}{4} + \frac{1}{4}e^{-4t}\right), \qquad t \in \mathbb{R}.$$

(c) Correlation of overlapping linear combinations. Let $v = \mathrm{Var}(R) > 0$. Independence gives $\mathrm{Cov}(R_1, R_2) = 0$. For $T = R_1 + R_2$,

$$\begin{aligned} \mathrm{Cov}(T, S) &= \mathrm{Cov}(R_1 + R_2, R_1 - 2R_2) = v - 2v = -v, \\ \mathrm{Var}(T) &= v + v = 2v, \qquad \mathrm{Var}(S) = v + 4v = 5v. \end{aligned}$$

The constant 3 in S affects neither covariance nor variance. Therefore

$$\mathrm{Corr}(T, S) = \frac{-v}{\sqrt{(2v)(5v)}} = -\frac{1}{\sqrt{10}}.$$

A numerical calculation is equally valid: the PMF gives $\mathbb{E}[R] = 0$ and $v = 3/2$.

Problem 4. Sessions with varying success rates

(a) Conditional variance. Conditional on $Q = q$, K is binomial with four trials and success probability q. Therefore

$$\mathrm{Var}(K \mid Q = q) = 4q(1 - q), \qquad q \in \{1/4, 3/4\}.$$

At either permitted value of q, this equals $3/4$.

(b) Unconditional variance. The conditional mean is $\mathbb{E}[K \mid Q] = 4Q$, taking values $1, 3$ with equal probabilities. Its mean is 2 and its variance is 1. Thus

$$\begin{aligned}\mathrm{Var}(K) &= \mathbb{E}[\mathrm{Var}(K \mid Q)] + \mathrm{Var}(\mathbb{E}[K \mid Q]) \\ &= \frac{3}{4} + \left\{\frac{1}{2}(1-2)^2 + \frac{1}{2}(3-2)^2\right\} = \frac{7}{4}.\end{aligned}$$

The variability between sessions' success rates contributes the second term.

(c) A sufficient number of sessions. Independent sessions give

$$\mathbb{E}[\overline{K}_n] = \mathbb{E}[K] = 2, \qquad \mathrm{Var}(\overline{K}_n) = \frac{\mathrm{Var}(K)}{n} = \frac{7}{4n}.$$

Chebyshev, applied directly at threshold $1/2$, yields

$$P\big(|\overline{K}_n - \mathbb{E}[K]| \geq 1/2\big) \leq \frac{7/(4n)}{(1/2)^2} = \frac{7}{n}.$$

The requested guarantee follows when

$$P\big(|\overline{K}_n - \mathbb{E}[K]| < 1/2\big) \geq 1 - \frac{7}{n} \geq 0.95, \quad \text{that is,} \quad n \geq 140.$$

Answer: $n = 140$, the smallest positive integer certified by this direct Chebyshev calculation. It need not be the smallest sample size for which the actual probability is at least 0.95.

Problem 5. Optional bonuses

(a) A sharp joint tail bound. Zero covariance and the means imply

$$\mathbb{E}[XY] = \mathrm{Cov}(X, Y) + \mathbb{E}[X]\mathbb{E}[Y] = 1.$$

On $\{X \geq 2, Y \geq 2\}$, the nonnegative product XY is at least 4. Hence Markov gives

$$P(X \geq 2, Y \geq 2) \leq P(XY \geq 4) \leq \frac{\mathbb{E}[XY]}{4} = \frac{1}{4}.$$

The following joint PMF attains this bound:

$$p_{X,Y}(x, y) = \begin{cases} 1/4, & (x, y) \in \{(0,0), (0,2), (2,0), (2,2)\}, \\ 0, & \text{otherwise.} \end{cases}$$

It has total mass 1, nonnegative support, and finite second moments. Directly, $\mathbb{E}[X] = \mathbb{E}[Y] = 1$ and $\mathbb{E}[XY] = 4(1/4) = 1$, so $\mathrm{Cov}(X, Y) = 0$. The target event occurs only at $(2, 2)$, with probability $1/4$. Thus the largest possible value is $1/4$. The upper-bound argument does not assume independence.

(b) Constant conditional means without independence. One valid joint PMF is

$$p_{X,Y}(x, y) = \begin{cases} 1/4, & (x, y) \in \{(0,-1), (0,1), (1,-2), (1,2)\}, \\ 0, & \text{otherwise.} \end{cases}$$

It sums to 1, and both variables are nonconstant. Each value of X has probability $1/2$, giving

$$\mathbb{E}[Y \mid X = 0] = \frac{-1+1}{2} = 0, \qquad \mathbb{E}[Y \mid X = 1] = \frac{-2+2}{2} = 0.$$

Nevertheless, $P(Y = 1 \mid X = 0) = 1/2 \neq 1/4 = P(Y = 1)$, so X, Y are dependent. Their conditional means agree, but their conditional distributions differ.