Department of Applied Mathematics and Statistics
The Johns Hopkins University

Introductory Examination–Winter Session

Tuesday, January 24, 2017

Instructions: Read carefully!

1. This **closed-book** examination consists of 15 problems, each worth 5 points. The passing grade has been set at 50 points, i.e., 2/3 of the total points. Partial credit will be given as appropriate; each part of a problem will be given the same weight. If you are unable to prove a result asserted in one part of a problem, you may still use that result to help in answering a later part.

2. You have been provided with a syllabus indicating the scope of the exam. Our purpose is to test not only your knowledge, but also your ability to apply that knowledge, and to provide mathematical arguments presented in **clear, logically justified steps**. The grading will reflect that broader purpose.

3. The problems have not been grouped by topic, but there are roughly equally many mainly motivated by each of the three areas identified in the syllabus (linear algebra; real analysis; probability). Nor have the problems been arranged systematically by difficulty. If a problem directs you to use a particular method of analysis, you *must* use it in order to receive substantial credit.

4. Start your answer to each problem on a NEW sheet of paper. Write only on ONE SIDE of each sheet, and please do not write very near the margins on any sheet. Arrange the sheets in order, and write your NAME and the PROBLEM NUMBER on each sheet.

5. The examination will begin at 8:30 AM; lunch and refreshments will be provided. The exam will end just before 5:00 PM. You may leave before then, but in that case you may not return.

6. Paper will be provided, but you should bring and use writing instruments that yield marks dark enough to be read easily.

7. **No calculators of any sort are needed or permitted.**

1. Prove that a continuous function $f : [0, 1] \to \mathbb{R}$ satisfies

$$\int_0^1 f(x)\psi(x)\, dx = \int_0^1 f(x)\, dx \int_0^1 \psi(x)\, dx$$

for all continuous functions $\psi : [0, 1] \to \mathbb{R}$ if and only if f is constant.

Solution: If $f(x) \equiv c$ is constant, then

$$\int_0^1 f(x)\psi(x)\, dx = c \int_0^1 \psi(x)\, dx = \int_0^1 f(x)\, dx \int_0^1 \psi(x)\, dx,$$

as desired. We present three proofs for the converse.

Proof #1: By continuity of f it suffices to prove that f is constant on $(0, 1)$. We will show that $f(x_0) = \int_0^1 f(x)\, dx$ for each $x_0 \in (0, 1)$.

Fix the value x_0. It is easy to construct a sequence of nonnegative continuous functions $\psi_n : [0, 1] \to \mathbb{R}$ vanishing outside the interval $(x_0 - n^{-1}, x_0 + n^{-1})$ such that $\int_0^1 \psi_n(x)\, dx = 1$. [For example, one can suitably center and scale the function $\phi : \mathbb{R} \to \mathbb{R}$ defined by

$$\phi(x) := \exp\left\{-\left(1 - 4\left|x - \tfrac{1}{2}\right|^2\right)^{-1}\right\} \mathbf{1}\,(0 < x < 1)\,.]$$

We then have

$$f(x_0) = \lim_{n\to\infty} \int_0^1 f(x)\psi_n(x)\, dx = \lim_{n\to\infty} \left[\int_0^1 f(x)\, dx \int_0^1 \psi_n(x)\, dx\right] = \int_0^1 f(x) dx,$$

where the first equality follows from the assumed continuity of f, the second from the stated assumption, and the third from the fact that each ψ_n integrates to 1.

Proof #2: We work in the Hilbert space $L^2[0, 1]$ and denote the function with constant value 1 by $\mathbf{1}$. Choosing $\psi = f$ gives the third equality in the following:

$$\|f\|^2\|\mathbf{1}\|^2 = \left[\int_0^1 f^2(x)\, dx\right]\left[\int_0^1 1\, dx\right] = \int_0^1 f^2(x)\, dx = \left[\int_0^1 f(x)\, dx\right]^2 = \langle f, \mathbf{1}\rangle^2.$$

Because we have inequality in the Cauchy–Schwarz inequality, it follows that $f \in L^2$ is a scalar multiple of $\mathbf{1}$, i.e., that f is almost surely constant. But f is continuous, so f is constant.

Proof #3: Assume the identity true for all continuous ψ. Take ψ such that $\psi(x) = x - a$ on $[0, a)$ and $\psi(x) = 0$ on $[a, 1]$, with $a \in (0, 1)$. The assumption implies that

$$\int_0^a f(x)(x-a)\,dx = -\frac{a^2}{2}\int_0^1 f\,dx.$$

Since f is continuous, the left-hand side is differentiable in a, and computing derivatives yields

$$-\int_0^a f(x)\,dx = -a\int_0^1 f(x)\,dx.$$

Taking another derivative implies

$$f(a) = \int_0^1 f(x)\,dx.$$

This is true for all $a \in (0, 1)$, and can be extended to $a = 0$ and $a = 1$ by continuity. This proves that f is constant.

---

2. Let A and B be two $n \times n$ real matrices. Show that if $AB = 0$ then

$$\operatorname{rank}(A) + \operatorname{rank}(B) \le n.$$

Solution: It follows from $AB = 0$ that $\operatorname{range}(B) \subseteq \operatorname{null}(A)$. Therefore,

$$\operatorname{rank}(A) + \operatorname{rank}(B) \le \dim(\operatorname{range}(A)) + \dim(\operatorname{null}(A)) = n.$$

---

3. Let A and B be $n \times m$ real matrices. Prove that a necessary and sufficient condition that there exists an $m \times m$ real matrix C such that $AC = B$ is that the column space of B is a subspace of the column space of A. (The *column space* of a matrix is the vector space spanned by its columns.)

Solution: Write A as $A = [\mathbf{a}_1\,\mathbf{a}_2\,\cdots\,\mathbf{a}_m]$ and B as $B = [\mathbf{b}_1\,\mathbf{b}_2\,\cdots\,\mathbf{b}_m]$. The column space of B is a subspace of the column space of A if and only if each $\mathbf{b}_j$ is a real linear combination of the columns of A, i.e., if and only if for each $j = 1, 2, \ldots, m$ there exist real c_{ij} $(i = 1, 2, \ldots, m)$ such that $\mathbf{b}_j = \sum_{i=1}^m \mathbf{a}_i c_{ij}$, i.e., if and only if there exists an $m \times m$ real matrix $C = [c_{ij}]$ such that $AC = B$.

4. Prove that there is a unique solution f of the integro-differential equation

$$f'(x) = f(x) + \int_0^1 f(y)\,dy, \quad x \in \mathbb{R},$$

with f twice-differentiable and $f(0) = 1$, and find that solution.

Solution: Suppose first that f is a solution. Differentiating both sides of the integro-differential equation gives

$$f''(x) = f'(x),$$

whose general solution is $f(x) = Ae^x + B$. Then

$$f'(x) = Ae^x = f(x) - B,$$

so that

$$-B = \int_0^1 f(y)\,dy = A(e-1) + B,$$

implying

$$B = -\frac{e-1}{2}A.$$

Furthermore,

$$f(0) = A + B = 1,$$

yielding

$$A = \frac{2}{3-e}, \quad B = -\frac{e-1}{3-e}$$

and thus finally

$$f(x) = \frac{2e^x - (e-1)}{3-e}.$$

Conversely, if f is defined by this formula, then f is twice-differentiable with $f(0) = 1$, and f satisfies the integro-differential equation.

5. Let $f : \mathbb{R} \to \mathbb{R}$ be a differentiable function that satisfies $f'(x) \leq \frac{1}{2}$ for all x.

   Prove that f has a unique fixed point, that is, that there exists one and only one real value a such that $f(a) = a$.

Solution: First we show that f has a fixed point a. To that end, let $g(x) := f(x) - x$. We seek a value a such that $g(a) = 0$. Suppose, for contradiction, that no such value exists. Then $g > 0$ or $g < 0$.

Note that $g'(x) = f'(x) - 1 \le -\frac{1}{2}$.

By the mean value theorem we have, for all $x \ne y$, that

$$\frac{g(x) - g(y)}{x - y} = g'(z) \le -\frac{1}{2},$$

where z is some value between x and y.

If $g > 0$, then for $x > 0$ we have

$$0 < \frac{g(x)}{x} = \frac{g(x) - g(0)}{x - 0} + \frac{g(0)}{x} \le -\frac{1}{2} + \frac{g(0)}{x}, \qquad (*)$$

which is clearly false for x sufficiently large.

If $g < 0$, then for $x < 0$ the result $(*)$ again holds, and this time we obtain a contradiction by considering $x \to -\infty$.

Therefore f has a fixed point a.

Note that f can have at most one fixed point because if b is another fixed point, then by the mean value theorem we have, for some value c between a and b, that

$$1 = \frac{a - b}{a - b} = \frac{f(a) - f(b)}{a - b} = f'(c) \le \frac{1}{2},$$

which is a contradiction.

---

6. Find the radius of convergence of the power series $\sum_{n=1}^{\infty} \frac{x^n}{n3^n}$, and then compute the value of $\sum_{n=1}^{\infty} \frac{1}{n3^n}$.

Solution: We use the root test: The key quantity $r := \lim_{n\to\infty} \sqrt[n]{\left|\frac{x^n}{n3^n}\right|}$ equals $\frac{|x|}{3}$, which shows that the series is divergent for $|x| > 3$ and absolutely convergent for $|x| < 3$. Therefore, **the radius of convergence is** 3.

Now let's define for $|x| < 3$ the function $f(x) := \sum_{n=1}^{\infty} \frac{x^n}{n3^n}$. We are asked to find $f(1)$. Clearly, $f(0) = 0$. Furthermore, recognizing a geometric series, we have

$$f'(x) = \sum_{n=1}^{\infty} \frac{x^{n-1}}{3^n} = \frac{1}{3}\sum_{n=1}^{\infty} \left(\frac{x}{3}\right)^{n-1} = \frac{1}{3}\left(\frac{1}{1 - \frac{x}{3}}\right) = \frac{1}{3 - x}.$$

Consequently, **the value of the series is**

$$f(1) = f(1) - f(0) = \int_0^1 \frac{1}{3 - x}\, dx = -\ln(3 - x)\Big|_{x=0}^{x=1} = \ln(3/2).$$

---

7. Show that if E and F are positive definite $n \times n$ matrices, then

$$\det(E+F) \geq \det E + \det F.$$

You may use without proof the existence of a positive definite square root $E^{1/2}$ of a positive definite matrix E.

Solution: The key is to use the multiplicativity of determinant. It then follows that

$$\det(E+F) = \det E \cdot \det(I + E^{-1}F)$$

and that $\det E > 0$ and

$$\det(I + E^{-1}F) = \det[E^{1/2}(I + E^{-1}F)E^{-1/2}] = \det(I + E^{-1/2}FE^{-1/2}).$$

But $E^{-1/2}FE^{-1/2}$ is clearly positive definite, and it is clear from consideration of eigenvalues that $\det(I + C) \geq 1 + \det C$ for any positive semidefinite matrix C. Putting the pieces together we find that

$$\begin{aligned} \det(E+F) &\geq (\det E)[1 + \det(E^{-1/2}FE^{-1/2})] = (\det E)\left(1 + \frac{\det F}{\det E}\right) \\ &= \det E + \det F, \end{aligned}$$

as desired.

---

8. Suppose X and Y are jointly continuous random variables having joint probability density function (pdf)

$$f(x,y) = \begin{cases} x+y, & \text{if } 0 < x < 1 \text{ and } 0 < y < 1; \\ 0, & \text{otherwise.} \end{cases}$$

Find the pdf f_W of $W := X + Y$.

Solution: It's easy to see that f is indeed a joint pdf, since it is nonnegative and integrates to 1. For $0 < w < 1$ we have

$$f_W(w) = \int_{-\infty}^{\infty} f(x, w-x)\,dx = \int_0^w f(x, w-x)\,dx = \int_0^w w\,dx = w^2.$$

For $1 \le w < 2$ we have

$$f_W(w) = \int_{-\infty}^{\infty} f(x, w-x)\, dx = \int_{w-1}^{1} f(x, w-x)\, dx = \int_{w-1}^{1} w\, dx = w(2-w).$$

For $w \le 0$ or $w \ge 2$ we have $f_W(w) = 0$. In summary, the (unique continuous) pdf of W is

$$f_W(w) = \begin{cases} w^2, & \text{if } 0 < w < 1 \\ w(2-w), & \text{if } 1 \le w < 2 \\ 0, & \text{otherwise} \end{cases}.$$

---

9. Let f_n:$[0,1] \to \mathbb{R}$, $n \in \mathbb{N}$, be a sequence of continuous functions and let f:$[0,1] \to \mathbb{R}$ be another continuous function. Show that f_n converges uniformly to f if and only if for every sequence $(x_n)_{n\in\mathbb{N}}$ such that x_n converges to some $x \in [0,1]$, we have $\lim_{n\to\infty} f_n(x_n) = f(x)$.

Solution: Let $\|\cdot\| \equiv \|\cdot\|_\infty$ denote sup-norm, and note that f_n converges uniformly to f if and only if $\|f_n - f\| \to 0$.

($\Rightarrow$) Assume that f_n converges uniformly to f and let $x_n \in [0,1]$, $n \in \mathbb{N}$, be such that $x_n \to x$ for some $x \in [0,1]$. Then

$$|f_n(x_n) - f(x)| \le |f_n(x_n) - f(x_n)| + |f(x_n) - f(x)| \le \|f_n - f\|_\infty + |f(x_n) - f(x)|.$$

In the limit as $n \to \infty$, the first term in the last bound vanishes by uniform convergence and the second vanishes by the continuity of f (at x).

($\Leftarrow$) Conversely, assume that for every sequence (x_n) from $[0,1]$ such that $x_n \to x$ for some $x \in [0,1]$ we have

$$\lim_{n\to\infty} f_n(x_n) = f(x).$$

Write $g_n := f_n - f$ for $n \in \mathbb{N}$. We will show that

$$\|g_n\| \to 0 \text{ along some subsequence.} \qquad (*)$$

Applying this result to each *subsequence* of (f_n), it then follows that every subsequence of $(\|g_n\|)$ has a further subsequence vanishing in the limit; hence $\|g_n\| \to 0$, as desired.

It remains to prove $(*)$. Since $|g_n|$ is continuous on the compact set $[0,1]$, there exists $y_n \in [0,1]$ such that

$$\|g_n\| = |g_n(y_n)|.$$

By the Bolzano–Weierstrass theorem, there exists a subsequence (y_{n_k}) of (y_n) such that $y_{n_k} \to x$ for some $x \in [0,1]$ as $k \to \infty$. Define

$$x_n := \begin{cases} y_n, & \text{if } n = n_k \text{ for some } k \in \mathbb{N}; \\ x, & \text{otherwise.} \end{cases}$$

Then $x_n \to x$, and so by assumption $f_n(x_n) \to f(x)$; further, $f(x_n) \to f(x)$ by the continuity of f. In particular,

$$\|g_{n_k}\| = |f_{n_k}(y_{n_k}) - f(y_{n_k})| = |f_{n_k}(x_{n_k}) - f(x_{n_k})| \to |f(x) - f(x)| = 0,$$

yielding $(*)$.

---

10. Take $N > 1$ to be a positive integer and $x_1, \ldots, x_N$ to be real numbers. Let $\mu := \frac{1}{N}\sum_{i=1}^N x_i$ and $\sigma^2 := \frac{1}{N}\sum_{i=1}^N (x_i - \mu)^2$. Suppose we draw I_1 uniformly from $\{1, \ldots, N\}$ and then, conditionally given $I_1 = i$, we draw I_2 uniformly from $\{1, \ldots, N\} \setminus \{i\}$. Define $X_i := x_{I_i}$ for $i = 1, 2$. Show that $\operatorname{Cov}(X_1, X_2) = -\sigma^2/(N-1)$.

Solution: We condition on I_1 and use the law of total covariance:

$$\operatorname{Cov}(X_1, X_2) = \operatorname{Cov}(\mathbb{E}(X_1|I_1), \mathbb{E}(X_2|I_1)) + \mathbb{E}\operatorname{Cov}(X_1, X_2 \,|\, I_1).$$

Observe first that the conditional distribution of X_1 given $I_1 = i_1$ is degenerate at the value x_{i_1}, and therefore $\mathbb{E}(X_1|I_1) = x_{I_1}$. Next, from the stated conditional distribution of I_2 given I_1 we see that

$$\mathbb{E}(X_2|I_1) = \frac{N\mu - x_{I_1}}{N-1}.$$

Therefore, the first contribution to $\operatorname{Cov}(X_1, X_2)$ is

$$\operatorname{Cov}(\mathbb{E}(X_1|I_1), \mathbb{E}(X_2|I_1)) = \operatorname{Cov}\left(x_{I_1}, \frac{N\mu - x_{I_1}}{N-1}\right) = -\frac{1}{N-1}\operatorname{Var} x_{I_1} = -\frac{\sigma^2}{N-1}.$$

Further,

$$\operatorname{Cov}(X_1, X_2 \,|\, I_1) = \operatorname{Cov}(x_{I_1}, X_2 \,|\, I_1) = 0,$$

so the second contribution $\mathbb{E}\operatorname{Cov}(X_1, X_2 \,|\, I_1)$ to $\operatorname{Cov}(X_1, X_2)$ vanishes. The desired result follows.

Solution #2: The joint probability mass function of I_1 and I_2 is given by

$$\mathbb{P}\{I_1 = i_1,\, I_2 = i_2\} = \begin{cases} 1/[N(N-1)], & \text{if } i_1 \neq i_2; \\ 0, & \text{if } i_1 = i_2, \end{cases}$$

and from this we see that I_1 and I_2 are each uniformly distributed in $\{1, \ldots, N\}$. Now

$$\mathbb{E}\, X_1 = \sum_{i=1}^{N} \mathbb{P}\{I_1 = i\} x_i = \sum_{i=1}^{N} \tfrac{1}{N} x_i = \mu,$$

and since I_2 has the same distribution as I_1 we have $\mathbb{E}\, X_2 = \mu$, as well. Further,

$$\begin{aligned}
\mathbb{E}[X_1 X_2] &= \frac{1}{N(N-1)} \sum_{1 \le i \ne j \le N} x_i x_j \\
&= \frac{1}{N(N-1)} \sum_{1 \le i,j \le N} x_i x_j - \frac{1}{N(N-1)} \sum_{i=1}^{N} x_i^2 \\
&= \frac{N}{N-1} \sum_{i,j=1}^{N} \frac{x_i x_j}{N^2} - \sum_{i=1}^{N} \frac{x_i^2}{N(N-1)}.
\end{aligned}$$

But

$$\sum_{i,j=1}^{N} \frac{x_i x_j}{N^2} = \mu^2$$

and

$$\sum_{i=1}^{N} \frac{x_i^2}{N(N-1)} = \frac{1}{N-1} \sum_{i=1}^{N} \frac{x_i^2}{N} = \frac{1}{N-1}(\sigma^2 + \mu^2).$$

So

$$\mathbb{E}[X_1 X_2] = \frac{N}{N-1}\mu^2 - \frac{1}{N-1}(\sigma^2 + \mu^2) = -\frac{1}{N-1}\sigma^2 + \mu^2,$$

and we conclude

$$\mathrm{Cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = -\frac{1}{N-1}\sigma^2.$$

---

11. Let X denote the number of different days of the year that are birthdays of four persons selected randomly. Calculate $\mathbb{E}\, X$. *You may assume that persons' birthdays are independent and uniformly distributed over 365 days of the year.*

Solution: We first give a simple solution. For $k = 1, 2, 3, \ldots, 365$, let I_k denote the indicator of the event that day k is the birthday of at least one of the persons. Then $X = I_1 + I_2 + \cdots + I_{365}$, and by linearity of expectation

$$\begin{aligned}
\mathbb{E}\, X &= \mathbb{E}[I_1 + I_2 + \cdots + I_{365}] = \mathbb{E}\, I_1 + \mathbb{E}\, I_2 + \cdots + \mathbb{E}\, I_{365} \\
&= 365\, \mathbb{E}\, I_1 = 365\, \mathbb{P}\{\text{day 1 is the birthday of at least one of the four people}\}.
\end{aligned}$$

By using complementation, this equals

$$365\,[1 - \mathbb{P}\{\text{day 1 is not the birthday of any of the four people}\}] = 365\left[1 - \left(\frac{364}{365}\right)^4\right],$$

which, if calculated, is 3.98359. . . .

Solution #2: Computation of $\mathbb{E}\,X$ by means of the probability mass function (call it p) for X is more difficult. Standard combinatorial arguments reveal that for an N-day year and r people we have

$$p(k) = N^{-r}\binom{N}{k}\sum_{\mathbf{c}\in C_{r,k}}\binom{r}{\mathbf{c}},$$

where the sum of k-nomial coefficients here is over the set $C_{r,k}$ of all k-part compositions $\mathbf{c} = (c_1, c_2, \ldots c_k)$ of the integer r, that is, over k-tuples of (strictly) positive integers summing to r. By the principle of inclusion–exclusion and the multinomial formula, this sum equals

$$\sum_{j=0}^{k-1}(-1)^j\binom{k}{j}(k-j)^r.$$

Thus

$$p(k) = N^{-r}\binom{N}{k}\sum_{j=0}^{k-1}(-1)^j\binom{k}{j}(k-j)^r.$$

From this second formula for $p(k)$ can be shown the general fact that the expectation $\sum_{k=1}^{r} k\,p(k)$ of X equals $N[1 - ((N-1)/N)^r]$. We give a direct computational proof in the case $r = 4$ (and $N = 365$) discussed in the problem using the second formula; the first formula can also be used for this small value of r.

In this case we have $p(k) = N^{-4}\binom{N}{k}s(k)$ where

$$\begin{aligned} s(1) &= 1, \\ s(2) &= 16 - 2 = 14, \\ s(3) &= 81 - 48 + 3 = 36, \\ s(4) &= 256 - 324 + 96 - 4 = 24. \end{aligned}$$

Thus

$$\mathbb{E}\,X = N^{-4}\left[\binom{N}{1} + 28\binom{N}{2} + 108\binom{N}{3} + 96\binom{N}{4}\right] = N\left[1 - \left(\frac{N-1}{N}\right)^4\right].$$

---

12. Prove that there do not exist a number $\epsilon > 0$ and a *real* matrix A that satisfy

$$A^{100} = \begin{bmatrix} -1 & 0 \\ 0 & -(1+\epsilon) \end{bmatrix}.$$

Solution: For a proof by contradiction, suppose that $\epsilon > 0$ and that A is a matrix with real entries that satisfies the equation [call it $(*)$]. Let $B := A^{50}$. Writing out the equation that B^2 equals the right side of $(*)$ entry by entry, we find that either B is a real diagonal matrix, which is clearly impossible, or B is a real matrix with $b_{22} = -b_{11}$ and determinant equal both to 1 and to $1 + \epsilon$, which is also clearly impossible.

Solution #2: For another proof by contradiction, suppose that $\epsilon > 0$ and that A is a matrix with real entries that satisfies $(*)$. We let a and b denote the eigenvalues of A. It follows from $(*)$ that one of the eigenvalues, say a, satisfies

$$a^{100} = -1,$$

which means that $a = x + iy \in \mathbb{C}$ with $y \neq 0$. Also, since the coefficients of the characteristic polynomial of A are functions of the *real* entries of A, they must also be real. Thus, the eigenvalues of A must come in complex conjugate pairs, so that

$$b = \bar{a} = x - iy.$$

Since a and b are conjugate, they have the same magnitude; this may be used in conjunction with $(*)$ to conclude that

$$1 = |a|^{100} = |b|^{100} = (1+\epsilon)^{100},$$

which is a contradiction. This completes the proof.

---

13. The ages of prospective married parents at a certain hospital can be approximated by a bivariate normal distribution with parameters $\mu_X = 28.2$, $\sigma_X = 6.0$, $\mu_Y = 31.5$, $\sigma_Y = 7.0$, and $\rho = 0.80$. (The parameters having label X refer to the pregnant women and those with label Y to the prospective father. The quantities μ are means and the quantities σ are standard deviations; ρ is the correlation.) For this hospital:

    (a) Consider the proportion of pregnant women who are over 30. Is this proportion closest to 10%, 40%, 60%, or 90%?

    (b) Consider the proportion of prospective fathers aged 35 who have wives over 30. Is this proportion closest to 10%, 40%, 60%, or 90%?

HINT: Recall—no proof need be given—that if (X, Y) has a bivariate normal distribution with means (μ_X, μ_Y), variances (σ_X^2, σ_Y^2), and correlation ρ, then the conditional distribution of X given $Y = y$ is normal with

$$\text{mean} \quad \mu_{X|Y=y} = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y) \quad \text{and} \quad \text{variance} \quad \sigma^2_{X|Y=y} = \sigma_X^2(1 - \rho^2).$$

You should not require a table of the normal distribution to solve this problem.

Solution:

(a) The desired proportion has the approximation

$$\begin{aligned}\mathbb{P}(X > 30) &= \mathbb{P}\left\{\frac{X - \mu_X}{\sigma_X} > \frac{30 - \mu_X}{\sigma_X}\right\} \\ &= \mathbb{P}\left\{Z > \frac{30 - 28.2}{6.0}\right\} = \mathbb{P}\left\{Z > \frac{1.8}{6.0}\right\} = \mathbb{P}\{Z > 0.3\} \doteq \mathbf{40}\%.\end{aligned}$$

(b) With $y = 35$, the conditional moments have approximate values

$$\begin{aligned}\mu_{X|Y=y} &= \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y) = 28.2 + (0.80)\frac{6.0}{7.0}(35 - 31.5) = 30.6, \\ \sigma^2_{X|Y=y} &= \sigma_X^2(1 - \rho^2) = (6.0)^2[1 - (0.80)^2] = (3.6)^2.\end{aligned}$$

Thus the desired proportion has the approximation

$$\mathbb{P}\{X > 30 \,|\, Y = 35\} \doteq \mathbb{P}\left\{Z > \frac{30 - 30.6}{3.6}\right\} = \mathbb{P}\{Z > -1/6\} \doteq \mathbf{60}\%.$$

---

14. A vector x in $\mathbb{R}^n$ has length 6. A vector y in $\mathbb{R}^n$ has the property that for every pair of real scalars a and b the vectors $ax + by$ and $4bx - 9ay$ are orthogonal. Compute the length of y and of $2x + 3y$.

Solution: For any scalars a and b we have

$$0 = \langle ax + by, 4bx - 9ay\rangle = 4ab\langle x, x\rangle - 9ab\langle y, y\rangle + (4b^2 - 9a^2)\langle x, y\rangle. \qquad (*)$$

In particular, choosing nonzero a and b in $(*)$ satisfying $4b^2 = 9a^2$ and using the fact that $\langle x, x\rangle = 36$ we obtain $\langle y, y\rangle = 16$, that is, **the length of y is** 4. Further, by taking $a = 1$ and $b = 0$ in $(*)$ we find that $\langle x, y\rangle = 0$. Therefore, $\langle 2x + 3y, 2x + 3y\rangle = 4\langle x, x\rangle + 9\langle y, y\rangle = 144 + 144 = 2 \cdot 144$, implying that **the length of $2x + 3y$ is** $12\sqrt{2}$.

15. Let X_1 and X_2 be independent normal random variables, each with mean zero but (perhaps) different variances ν_1 and ν_2. Correspondingly, we write the probability density function for X_j as $\varphi_{X_j}(\cdot) = \varphi(\cdot; 0, \nu_j)$. Derive the maximally-simplified closed-form expression for

$$\mathbb{E}[X_1 | X_1 + X_2 = s]$$

in terms of just s, ν_1, and ν_2.

Solution: Let $S := X_1 + X_2$ and $\nu := \nu_1 + \nu_2$. Because X_1 and X_2 are independent normal random variables, their distribution is bivariate normal, with mean vector zero and diagonal covariance matrix $\mathrm{diag}[\nu_1, \nu_2]$. Thus S and X_1 are bivariate normal with mean vector zero, variances ν and ν_1, respectively, and covariance ν_1. It is well known that, when T and U are jointly normal random vectors with respective mean vectors μ_T and μ_U, respective covariance matrices Σ_{TT} and Σ_{UU}, and cross-covariance matrix $\Sigma_{TU} = \Sigma_{UT}^T$, the conditional distribution of U given $T = t$ is normal with mean vector $\mu_U + \Sigma_{UT}\Sigma_{TT}^{-1}(t - \mu_T)$ and covariance matrix $\Sigma_{UU.T} := \Sigma_{UU} - \Sigma_{UT}\Sigma_{TT}^{-1}\Sigma_{TU}$. (In the bivariate case pertinent here, this fact is reviewed in the hint to Problem 13!) In our case we therefore see that the conditional distribution of X_1 given $S = s$ is normal with mean $(\nu_1/\nu)s$ and variance $\nu_1 - (\nu_1/\nu)\nu_1 = \nu_1\nu_2/\nu$. In particular,

$$\mathbb{E}[X_1 | X_1 + X_2 = s] = \nu_1 s/(\nu_1 + \nu_2).$$

Solution #2: Let $S := X_1 + X_2$ and $\nu := \nu_1 + \nu_2$. By independence, the joint density of X_1 and X_2 is the product of their marginal densities. By change of variables, the joint density $\varphi_{X_1,S}$ of X_1 and S is therefore given by

$$\varphi_{X_1,S}(x, s) = \varphi_{X_1}(x)\varphi_{X_2}(s - x).$$

Further, the marginal distribution of S is normal with mean zero and variance ν, and hence the conditional density $\phi_{X_1|S}$ of X_1 given S is given by

$$\phi_{X_1|S}(x|s) = \frac{\phi_{X_1,S}(x, s)}{\varphi_\nu(s)} = \frac{\varphi_{X_1}(x)\varphi_{X_2}(s - x)}{\varphi_\nu(s)}$$

A small "complete-the-square" computation reveals that the conditional distribution of X_1 given S is normal with mean $\nu_1 s/\nu$ and variance $\nu_1\nu_2/\nu$. In particular,

$$\mathbb{E}[X_1 | X_1 + X_2 = s] = \nu_1 s/(\nu_1 + \nu_2).$$