The normal distribution and z-scores

Ten problems on the normal distribution: standardizing and reading Φ with its symmetry, the 68-95-99.7 rule for an initialization scale, a classifier threshold from an inverse z-score and the true positive rate it gives, z-scores that compare results on different scales, linear transforms, sums and differences of independent normals, the density's peak and inflection points, the normal approximation to a binomial with a continuity correction, a two-component Gaussian mixture, and the false alarms of a |z| > 3 outlier rule, with worked solutions and the mistakes of using a two-sided cutoff for a one-sided rate, adding standard deviations, dropping the continuity correction, reading a density as a probability and treating a mixture as one normal.

BEFORE YOU START

The normal distribution is the bell curve: the shape of measurement noise, of averages by the central limit theorem, and of the random weights a network starts from. Every normal is a shifted and stretched copy of one standard normal, so a single table, or a single function Φ, answers every question about all of them once a value is converted to a z-score. These ten problems practice that conversion in both directions, from a value to a probability and from a probability back to a value, then use the closure of normals under linear maps and independent sums, look at the shape of the density itself, approximate a binomial count, and end with two places where the normal is only a model: a mixture of two normals, and an outlier rule run on many points at once. The five mistakes at the end are the ones that give a believable number: a two-sided cutoff used for a one-sided rate, standard deviations added instead of variances, a count approximated without the continuity correction, a density value read as a probability, and a mixture summarized by one normal with the same mean and variance.

- **Normal**(μ, σ^2) has density $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$, mean μ and variance σ^2. The second parameter is the variance; the standard deviation is σ.
- **Standardizing.** If X is Normal(μ, σ^2), then $Z = (X - \mu)/\sigma$ is Normal$(0, 1)$. The value $z = (x - \mu)/\sigma$ is the z-score of x: how many standard deviations it lies from the mean.
- Φ is the standard normal CDF, $\Phi(z) = P(Z \le z)$, and $\varphi(z) = e^{-z^2/2}/\sqrt{2\pi}$ is its density. By symmetry $\Phi(-z) = 1 - \Phi(z)$. So $P(X \le x) = \Phi\big((x - \mu)/\sigma\big)$.
- **Values used here.** $\Phi(1) = 0.8413$, $\Phi(1.5) = 0.9332$, $\Phi(2) = 0.9772$, $\Phi(3) = 0.99865$. The cutoffs $z_{0.95} = 1.6449$ and $z_{0.975} = 1.9600$ satisfy $\Phi(z_q) = q$. Other values come from a table or `scipy.stats.norm.cdf` and `norm.ppf`.

- **Closure.** For constants $a \neq 0$ and b, $aX + b$ is Normal$(a\mu + b, a^2\sigma^2)$. For independent normals X_1, X_2, $X_1 \pm X_2$ is normal with mean $\mu_1 \pm \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$: variances add, in a sum or a difference.
- **Binomial.** A Binomial(n, p) count has mean np and variance $np(1-p)$. For large n it is close to a normal with those moments; since the count takes only whole values, $P(X \geq k)$ is approximated by $P(Y > k - \frac{1}{2})$, the continuity correction.

Builds on: Continuous random variables: densities, CDFs, expectation and the uniform, exponential and normal

Problems

1. A model's response time is Normal$(120, 20^2)$ in milliseconds. Find $P(X < 150)$, $P(X > 90)$ and $P(100 < X < 160)$. `difficulty ·`

2. A layer's weights are initialized from Normal$(0, 0.02^2)$. Use the 68-95-99.7 rule to estimate the fraction of weights with $|w| > 0.04$, the fraction in $(0.02, 0.04)$, and an interval that holds 99.7% of them. Then give the exact values from Φ. `difficulty ·`

3. A detector's score on negative examples is Normal$(0.3, 0.1^2)$, and on positive examples Normal$(0.7, 0.15^2)$. Choose the threshold t that flags 5% of negatives (the false positive rate), and find the fraction of positives it catches (the true positive rate). `difficulty ··`

4. A model scores 78 on benchmark A, where published models have mean 70 and standard deviation 5, and 62 on benchmark B, where they have mean 50 and standard deviation 10. Assuming both are roughly normal, on which benchmark is the result more exceptional? What fraction of models does it beat on each? `difficulty ·`

5. Show that if X is Normal(μ, σ^2) and $a > 0$, then $Y = aX + b$ is Normal$(a\mu + b, a^2\sigma^2)$, by computing its CDF. What changes for $a < 0$? Then a sensor reads temperature C as Normal$(20, 5^2)$ in degrees Celsius: find the distribution of $F = 1.8C + 32$ and $P(F > 77)$. `difficulty ··`

6. A request passes through preprocessing, Normal$(30, 4^2)$ ms, then inference, Normal$(80, 3^2)$ ms, independently. Find the distribution of the total time and $P(\text{total} > 120)$. Then find the distribution of inference minus preprocessing and the probability that the difference is under 40 ms. difficulty ··

7. For $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$, find the maximum value of f and where it occurs, and show that the inflection points are at $x = \mu \pm \sigma$. Evaluate the peak for $\sigma = 0.1$, and the ratio $f(\mu + \sigma)/f(\mu)$. difficulty ···

8. A classifier that guesses at random answers 100 balanced yes-or-no questions, so its number correct X is Binomial$(100, 0.5)$. Approximate $P(X \geq 60)$ with a normal and the continuity correction, and compare with the exact value 0.0284 and with the approximation without the correction. difficulty ··

9. A model's score is drawn from a Gaussian mixture: with probability 0.7 from Normal$(0, 1)$ and with probability 0.3 from Normal$(3, 1)$. Find its mean, its variance and $P(X > 4)$. Compare $P(X > 4)$ with a single normal of the same mean and variance. Then find the probability that a score of $x = 2$ came from the second component. difficulty ···

10. An anomaly detector flags any point with $|z| > 3$. On 10 000 points drawn independently from a normal distribution with known mean and standard deviation, find the probability that one point is flagged, the expected number of flags, and the probability of at least one. What threshold on $|z|$ gives one expected flag? difficulty ···

Answers

1. $P(X < 150) = 0.9332$; $P(X > 90) = 0.9332$; $P(100 < X < 160) = 0.8186$
2. About 5% (exactly 4.55%) have $|w| > 0.04$; about 13.5% (exactly 13.59%) lie in $(0.02, 0.04)$; 99.7% lie in $(-0.06, 0.06)$
3. $t \approx 0.4645$; true positive rate ≈ 0.9418
4. More exceptional on A ($z = 1.6$, beating about 94.5%) than on B ($z = 1.2$, beating about 88.5%)
5. Y is Normal$(a\mu + b, a^2\sigma^2)$ for every $a \neq 0$, with standard deviation $|a|\sigma$; F is Normal$(68, 9^2)$ and $P(F > 77) = 0.1587$
6. Total: Normal$(110, 5^2)$, $P(\text{total} > 120) = 0.0228$; difference: Normal$(50, 5^2)$, $P(D < 40) = 0.0228$
7. Maximum $1/(\sigma\sqrt{2\pi})$ at $x = \mu$, which is 3.989 for $\sigma = 0.1$; inflection points at $\mu \pm \sigma$, where the density is $e^{-1/2} \approx 0.607$ of its peak
8. $P(X \geq 60) \approx 0.0287$ with the correction, against 0.0284 exactly; 0.0228 without it
9. Mean 0.9, variance 2.89, $P(X > 4) \approx 0.0476$ (one normal with those moments gives 0.0341); a score of 2 came from the second component with probability ≈ 0.658
10. $P(\text{flag}) = 0.0027$; 27 expected flags and at least one almost surely, with no anomaly present; $|z| > 3.89$ gives one expected flag

Worked solutions

Problem 1

> A model's response time is Normal$(120, 20^2)$ in milliseconds. Find $P(X < 150)$, $P(X > 90)$ and $P(100 < X < 160)$.

$z = (150 - 120)/20 = 1.5$, so $P(X < 150) = \Phi(1.5) = 0.9332$	Standardize, then read Φ.
$z = (90 - 120)/20 = -1.5$	90 ms is 1.5 standard deviations below the mean.
$P(X > 90) = 1 - \Phi(-1.5) = \Phi(1.5) = 0.9332$	Symmetry: the area above -1.5 equals the area below 1.5.
$z = -1$ at 100 and $z = 2$ at 160	The two ends of the interval.
$\Phi(2) - \Phi(-1) = 0.9772 - (1 - 0.8413)$	The area below the top end minus the area below the bottom end.
$P(X < 150) = 0.9332$; $P(X > 90) = 0.9332$; $P(100 < X < 160) = 0.8186$	The first two are equal because 90 and 150 sit symmetrically about the mean. Whether $<$ or $\le$ is written makes no difference for a continuous variable.

Problem 2

> A layer's weights are initialized from Normal$(0, 0.02^2)$. Use the 68-95-99.7 rule to estimate the fraction of weights with $|w| > 0.04$, the fraction in $(0.02, 0.04)$, and an interval that holds 99.7% of them. Then give the exact values from Φ.

$\lvert w\rvert > 0.04$ means $\lvert z\rvert > 2$, and about 95% lies within 2σ, so about 5% lies outside	The rule: 68%, 95% and 99.7% within 1, 2 and 3 standard deviations.
$(0.02, 0.04)$ is $1 < z < 2$: about $(95\% - 68\%)/2 = 13.5\%$	The band between 1σ and 2σ has two equal halves, one on each side.

99.7% is within $3\sigma = 0.06$: the interval $(-0.06, 0.06)$	Three standard deviations either side.
Exact: $2(1 - \Phi(2)) = 0.0455$ and $\Phi(2) - \Phi(1) = 0.1359$	From the table values.
About 5% **(exactly** 4.55%**) have** $\lvert w \rvert > 0.04$; **about** 13.5% **(exactly** 13.59%**) lie in** $(0.02, 0.04)$; 99.7% **lie in** $(-0.06, 0.06)$	The rule is a rounded form of Φ. TensorFlow's truncated-normal initializer redraws every weight beyond 2σ, which here is about one weight in 22.

Problem 3

> A detector's score on negative examples is Normal$(0.3, 0.1^2)$, and on positive examples Normal $(0.7, 0.15^2)$. Choose the threshold t that flags 5% of negatives (the false positive rate), and find the fraction of positives it catches (the true positive rate).

$P(\text{negative score} > t) = 0.05$ means $\Phi\big((t - 0.3)/0.1\big) = 0.95$	A one-sided rate: all 5% sits in the upper tail.
$(t - 0.3)/0.1 = z_{0.95} = 1.6449$	Invert Φ.
$t = 0.3 + 1.6449 \cdot 0.1 = 0.4645$	Undo the standardization: $x = \mu + z\sigma$.
For positives, $z = (0.4645 - 0.7)/0.15 = -1.570$	Standardize the same threshold with the positives' mean and standard deviation.
$P(\text{positive score} > t) = 1 - \Phi(-1.570) = \Phi(1.570)$	Symmetry.
$t \approx 0.4645$; **true positive rate** ≈ 0.9418	About 5.8% of positives are missed. Moving t trades one error for the other, and tracing that trade as t varies draws the ROC curve.

Problem 4

A model scores 78 on benchmark A, where published models have mean 70 and standard deviation 5, and 62 on benchmark B, where they have mean 50 and standard deviation 10. Assuming both are roughly normal, on which benchmark is the result more exceptional? What fraction of models does it beat on each?

On A: $z = (78 - 70)/5 = 1.6$	Standard deviations above that benchmark's mean.
On B: $z = (62 - 50)/10 = 1.2$	B's raw gain of 12 points is larger, but B's scores are twice as spread out.
$\Phi(1.6) = 0.9452$ and $\Phi(1.2) = 0.8849$	The fraction of models scoring below each result.
More exceptional on A ($z = 1.6$, beating about 94.5%) than on B ($z = 1.2$, beating about 88.5%)	A z-score removes the units and the spread, so results on different scales can be compared. The comparison relies on both distributions having the same shape; the percentiles are only as good as the normal model.

Problem 5

Show that if X is Normal(μ, σ^2) and $a > 0$, then $Y = aX + b$ is Normal$(a\mu + b, a^2\sigma^2)$, by computing its CDF. What changes for $a < 0$? Then a sensor reads temperature C as Normal $(20, 5^2)$ in degrees Celsius: find the distribution of $F = 1.8C + 32$ and $P(F > 77)$.

$P(Y \le y) = P\big(X \le (y - b)/a\big)$	Solve $aX + b \le y$ for X; dividing by $a > 0$ keeps the inequality.
$= \Phi\Big(\dfrac{(y - b)/a - \mu}{\sigma}\Big)$	Standardize X.
$= \Phi\Big(\dfrac{y - (a\mu + b)}{a\sigma}\Big)$	Multiply the top and bottom by a.
That is the CDF of Normal$(a\mu + b, a^2\sigma^2)$	Step 3 is Φ of a z-score with mean $a\mu + b$ and standard deviation $a\sigma$.
For $a < 0$: $P(Y \le y) = P\big(X \ge (y - b)/a\big)$	Dividing by $a < 0$ flips the inequality.

$= 1 - \Phi\left(\frac{y-(a\mu+b)}{a\sigma}\right)$	Standardize, and multiply the top and bottom by a as in step 3.
$= \Phi\left(\frac{y-(a\mu+b)}{\lvert a\rvert\sigma}\right)$	Symmetry, $1-\Phi(u)=\Phi(-u)$, with $-a=\lvert a\rvert$: again a normal CDF, now with standard deviation $\lvert a\rvert\sigma$.
F has mean $1.8 \cdot 20 + 32 = 68$ and standard deviation $1.8 \cdot 5 = 9$	The closure rule with $a = 1.8$, $b = 32$.
$P(F > 77) = 1 - \Phi\big((77-68)/9\big) = 1 - \Phi(1)$	77 degrees Fahrenheit is 25 degrees Celsius, one standard deviation up on either scale.
Y is Normal$(a\mu+b, a^2\sigma^2)$ for every $a \neq 0$, with standard deviation $\lvert a\rvert\sigma$; F is Normal$(68, 9^2)$ and $P(F > 77) = 0.1587$	A change of units moves and stretches the distribution but leaves every z-score where it was. Standardizing a feature is the case $a = 1/\sigma$, $b = -\mu/\sigma$.

Problem 6

> A request passes through preprocessing, Normal$(30, 4^2)$ ms, then inference, Normal$(80, 3^2)$ ms, independently. Find the distribution of the total time and $P(\text{total} > 120)$. Then find the distribution of inference minus preprocessing and the probability that the difference is under 40 ms.

Mean of the total: $30 + 80 = 110$	Means add.
Variance of the total: $4^2 + 3^2 = 25$, so the standard deviation is 5	For independent variables, variances add.
$P(\text{total} > 120) = 1 - \Phi\big((120 - 110)/5\big) = 1 - \Phi(2)$	A sum of independent normals is normal.
The difference has mean $80 - 30 = 50$ and variance $3^2 + 4^2 = 25$	Subtracting is adding -1 times a variable, and $(-1)^2 = 1$: the variances still add.

$P(D < 40) = \Phi\big((40-50)/5\big) = \Phi(-2) = 1 - \Phi(2)$	Standardize the difference.
Total: Normal$(110, 5^2)$, $P(\text{total} > 120) = 0.0228$; **difference: Normal** $(50, 5^2)$, $P(D < 40) = 0.0228$	The standard deviation of the total is 5, not $4 + 3 = 7$ (Mistake 2). Independence is needed for the variances to add; correlated stages add a covariance term.

Problem 7

> For $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$, find the maximum value of f and where it occurs, and show that the inflection points are at $x = \mu \pm \sigma$. Evaluate the peak for $\sigma = 0.1$, and the ratio $f(\mu + \sigma)/f(\mu)$.

Let $u = (x - \mu)/\sigma$, so $f = \dfrac{1}{\sigma}\varphi(u)$ with $du/dx = 1/\sigma$	Every normal density is the standard one, moved and rescaled.
$\varphi'(u) = -u\,\varphi(u)$	Chain rule on $e^{-u^2/2}$.
$\varphi''(u) = (u^2 - 1)\,\varphi(u)$	Product rule: $-\varphi(u) - u\varphi'(u) = -\varphi + u^2\varphi$.
$f'(x) = 0$ only at $u = 0$, that is $x = \mu$, a maximum	$\varphi > 0$, and $-u$ changes sign from $+$ to $-$ there.
f'' changes sign where $u^2 = 1$: $x = \mu \pm \sigma$	$f''(x) = \varphi''(u)/\sigma^3$, and $\varphi > 0$, so the sign is that of $u^2 - 1$.
Peak $f(\mu) = 1/(\sigma\sqrt{2\pi})$; for $\sigma = 0.1$ it is 3.989	$1/\sqrt{2\pi} = 0.3989$.
$f(\mu + \sigma)/f(\mu) = e^{-1/2} = 0.6065$	$u = 1$ in the exponent.

Maximum $1/(\sigma\sqrt{2\pi})$ at $x = \mu$, which is 3.989 for $\sigma = 0.1$; inflection points at $\mu \pm \sigma$, where the density is $e^{-1/2} \approx 0.607$ of its peak	The inflection points are where the curve turns from bending down to bending up, which is the way to read σ off a plotted bell curve. A density above 1 is fine: it is probability per unit length, not a probability (Mistake 4).

Problem 8

> A classifier that guesses at random answers 100 balanced yes-or-no questions, so its number correct X is Binomial$(100, 0.5)$. Approximate $P(X \geq 60)$ with a normal and the continuity correction, and compare with the exact value 0.0284 and with the approximation without the correction.

Mean $np = 50$, variance $np(1-p) = 25$, standard deviation 5	The binomial's moments.
$P(X \geq 60) \approx P(Y > 59.5)$	The count 60 stands for the interval $(59.5, 60.5)$, so the area starts half a unit below 60.
$z = (59.5 - 50)/5 = 1.9$	Standardize.
$1 - \Phi(1.9) = 0.0287$	From a table.
Without the correction: $1 - \Phi\big((60 - 50)/5\big) = 1 - \Phi(2) = 0.0228$	This leaves out half of the bar at 60.
$P(X \geq 60) \approx 0.0287$ with the correction, against 0.0284 exactly; 0.0228 without it	The corrected value is off by 1% of the answer, the uncorrected one by 20%. A model that gets 60 of 100 right is unlikely, but not impossible, to be guessing: this is the p-value of a one-sided test.

Problem 9

> A model's score is drawn from a Gaussian mixture: with probability 0.7 from Normal$(0, 1)$ and with probability 0.3 from Normal$(3, 1)$. Find its mean, its variance and $P(X > 4)$. Compare

$P(X > 4)$ with a single normal of the same mean and variance. Then find the probability that a score of $x = 2$ came from the second component.

$E[X] = 0.7 \cdot 0 + 0.3 \cdot 3 = 0.9$	The mean of a mixture is the weighted mean of the means.
$E[X^2] = 0.7(1 + 0) + 0.3(1 + 9) = 3.7$	For each component, $E[X^2] = \sigma^2 + \mu^2$.
$\mathrm{Var}(X) = 3.7 - 0.9^2 = 2.89$, standard deviation 1.7	Within-component variance 1 plus the spread of the means, $0.7 \cdot 0.3 \cdot 3^2 = 1.89$.
$P(X > 4) = 0.7(1 - \Phi(4)) + 0.3(1 - \Phi(1))$	Total probability over the two components.
$= 0.7 \cdot 0.00003 + 0.3 \cdot 0.1587 = 0.0476$	The first component contributes almost nothing.
One normal: $1 - \Phi\big((4 - 0.9)/1.7\big) = 1 - \Phi(1.824) = 0.0341$	Same mean and variance, different shape.
Responsibility: $\dfrac{0.3\,\varphi(2 - 3)}{0.7\,\varphi(2) + 0.3\,\varphi(2 - 3)}$	Bayes' rule, with each component's density at $x = 2$.
$= \dfrac{0.3 \cdot 0.2420}{0.7 \cdot 0.0540 + 0.3 \cdot 0.2420} = \dfrac{0.0726}{0.1104}$	$\varphi(1) = \varphi(-1) = 0.2420$ and $\varphi(2) = 0.0540$.
Mean 0.9, **variance** 2.89, $P(X > 4) \approx 0.0476$ **(one normal with those moments gives** 0.0341**); a score of** 2 **came from the second component with probability** ≈ 0.658	The single normal underestimates this tail by about 28% (Mistake 5). The responsibility in step 7 is the E step of the EM algorithm for fitting a mixture.

Problem 10

An anomaly detector flags any point with $|z| > 3$. On $10\,000$ points drawn independently from a normal distribution with known mean and standard deviation, find the probability that one

> point is flagged, the expected number of flags, and the probability of at least one. What threshold on $|z|$ gives one expected flag?

$P(\|Z\| > 3) = 2(1 - \Phi(3)) = 0.0027$	Both tails.
Expected flags: $10\,000 \cdot 0.0027 = 27$	Linearity of expectation; independence is not needed for this.
$P(\text{at least one}) = 1 - (1 - 0.0027)^{10\,000}$	Independence: no point is flagged with probability 0.9973 each time.
$\approx 1 - e^{-27}$, which is 1 to eleven decimal places	$(1 - q)^n \approx e^{-nq}$ for small q.
One expected flag needs $2(1 - \Phi(c)) = 10^{-4}$, so $\Phi(c) = 0.99995$	Divide the target count by 10 000, and split it over two tails.
$P(\text{flag}) = 0.0027$; 27 **expected flags and at least one almost surely, with no anomaly present;** $\|z\| > 3.89$ **gives one expected flag**	A cutoff that is rare for one point is routine across thousands, the multiple-comparisons problem in another form. Real data usually have heavier tails than the normal, so the true count of false flags is often higher still.

Where this goes wrong

1. Using the two-sided cutoff for a one-sided rate

The number 1.96 is the one remembered alongside 5%, from confidence intervals and two-sided tests.

The threshold must flag 5% of negatives, Normal$(0.3, 0.1^2)$	Right so far: Problem 3.
"5% means $z = 1.96$."	The habit that causes the mistake: 1.96 leaves 2.5% in each of two tails. A false positive rate counts only the upper tail, so the cutoff is $z_{0.95} = 1.6449$.

$t = 0.3 + 1.96 \cdot 0.1 = 0.496$ ✗	This flags only 2.5% of negatives, and the true positive rate falls from 0.9418 to $\Phi(1.36) \approx 0.9131$. The detector is stricter than asked, and catches about 3 fewer positives in every 100.

2. Adding standard deviations

The means of the two stages add, and the standard deviation looks like the same kind of quantity.

The total has mean $30 + 80 = 110$ ms	Right so far: Problem 6.
"Its standard deviation is $4 + 3 = 7$ ms."	Variances add for independent variables, not standard deviations: $\sqrt{4^2 + 3^2} = 5$.
$P(\text{total} > 120) = 1 - \Phi(10/7) \approx 0.0766$ ✗	The right answer is $1 - \Phi(2) \approx 0.0228$, so the wrong spread makes a slow request look more than three times as common. Adding standard deviations is right only when the two times are perfectly correlated.

3. Dropping the continuity correction

A normal is continuous, so it seems natural to put the boundary of the count where it is written.

X is Binomial$(100, 0.5)$, approximated by Y, Normal$(50, 5^2)$	Right so far: Problem 8.
"$P(X \geq 60) \approx P(Y \geq 60)$."	Each whole count k stands for the interval $(k - \frac{1}{2}, k + \frac{1}{2})$ under the curve, so the bar at 60 starts at 59.5.
$P(X \geq 60) \approx 1 - \Phi(2) = 0.0228$ ✗	The exact value is 0.0284 and the corrected approximation 0.0287. Shifting the wrong way, to 60.5, is worse still: $1 - \Phi(2.1) \approx 0.0179$. For $\geq k$ use $k - \frac{1}{2}$; for $\leq k$ use $k + \frac{1}{2}$.

4. Reading a density value as a probability

The density is the function that gets plotted and evaluated, and its value at a point looks like the chance of that point.

X is Normal$(0, 0.1^2)$, with $f(0) = 3.989$	Right so far: Problem 7.
"$f(0)$ is the probability that X is 0."	A probability is an area under f; a single point has probability 0, and f is probability per unit length.
$P(X = 0) = 3.989$ ✗	A probability above 1 is impossible, which is the clue. The chance of landing within 0.01 of 0 is $2\Phi(0.1) - 1 \approx 0.0797$, close to $f(0) \cdot 0.02 \approx 0.0798$: density times width, for a short interval.

5. Treating a mixture as one normal

Any distribution has a mean and a variance, and a normal with the same two numbers looks like a fair summary of it.

The mixture of Problem 9 has mean 0.9 and standard deviation 1.7	Right so far.
"So X is approximately Normal$(0.9, 1.7^2)$."	A normal is determined by two numbers, but a mixture is not: this one has two bumps, and its upper tail is the tail of the second component.
$P(X > 4) \approx 1 - \Phi(1.824) \approx 0.0341$ ✗	The mixture gives 0.0476, about 40% more. A z-score is only a probability when the distribution really is normal; with two bumps it can be far off in either direction.

More at eulerpad.app. This is one of 82 free practice sets with worked solutions at eulerpad.app/practice, alongside guides to the topics behind them.

EulerPad, the app, reads your handwritten math on iPad, iPhone and Android and checks it line by line. It's free to start.

This set is also Chapter 12 of *Probability & Statistics for Machine Learning*, a 250-problem workbook with worked solutions: eulerpad.app/workbooks

eulerpad.app/practice/normal-distribution-and-z-scores