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

Ten problems on continuous random variables: normalizing a density and integrating it for a probability, a CDF turned into a density, a median and a mean with an infinite variance, expectation and variance by integration, the uniform distribution and a density above 1, the exponential's CDF, mean, median and memorylessness, standardizing a normal with z-values and a percentile, the density of Y = X² through its CDF, inverse-CDF sampling and the minimum of independent exponentials, with worked solutions and the mistakes of reading a density as a probability, dropping the |dx/dy| in a change of variables, taking the exponential's mean as λ, dividing by σ² and taking the mean as the median.

BEFORE YOU START

A continuous random variable takes values in an interval: a waiting time, a measurement, the output of a regression model. No single value has positive probability, so its distribution is described by a density $f(x)$, and probabilities are areas under it. Everything from the discrete case carries over with sums replaced by integrals: $E[X] = \int x f(x)\,dx$ instead of $\sum x\,p(x)$. What does not carry over is reading $f(x)$ as a probability. A density can be larger than 1; only its integral over an interval is a probability. These ten problems normalize densities, move between the density and the cumulative distribution function, work through the uniform, exponential and normal distributions, and end with the two tools behind sampling: transforming a variable through its CDF, and the inverse CDF that turns a uniform random number into any distribution.

- A **density** f has $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)\,dx = 1$, and $P(a < X < b) = \int_a^b f(x)\,dx$. In particular $P(X = a) = 0$, so $<$ and $\leq$ give the same probabilities.
- The **CDF** is $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt$. It rises from 0 to 1, $P(a < X \leq b) = F(b) - F(a)$, and $f = F'$ wherever F is differentiable.
- **Expectation.** $E[X] = \int x f(x)\,dx$ and $E[g(X)] = \int g(x) f(x)\,dx$; $\mathrm{Var}(X) = E[X^2] - (E[X])^2$. The median m solves $F(m) = \frac{1}{2}$.
- **Uniform**(a, b): $f(x) = \frac{1}{b-a}$ on (a, b), mean $\frac{a+b}{2}$, variance $\frac{(b-a)^2}{12}$.
- **Exponential**(λ): $f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$, $F(x) = 1 - e^{-\lambda x}$, $P(X > x) = e^{-\lambda x}$, mean $\frac{1}{\lambda}$, variance $\frac{1}{\lambda^2}$. λ is a rate, events per unit time; $\frac{1}{\lambda}$ is the average wait.
- **Normal**(μ, σ^2): $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$. Standardize with $Z = \frac{X-\mu}{\sigma}$, which is Normal $(0, 1)$ with CDF Φ. Useful values: $\Phi(1) = 0.8413$, $\Phi(2) = 0.9772$, $\Phi(1.2816) = 0.9$, and $\Phi(-z) = 1 - \Phi(z)$.

- **Change of variables.** For $Y = g(X)$, find $F_Y(y) = P(g(X) \le y)$ as a probability about X, then differentiate. For an increasing g this gives $f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$ with $x = g^{-1}(y)$.

Problems

1. X has density $f(x) = cx^2$ for $0 \le x \le 3$ and 0 elsewhere. Find c, the CDF, and $P(1 < X < 2)$. difficulty ·

2. X has CDF $F(x) = 1 - \dfrac{1}{x^2}$ for $x \ge 1$ and $F(x) = 0$ for $x < 1$. Find its density, $P(2 < X < 4)$, the median, $E[X]$ and $E[X^2]$. difficulty ·

3. X has density $f(x) = 2x$ on $[0, 1]$. Find $E[X]$, $E[X^2]$, $\mathrm{Var}(X)$, the standard deviation and the median. difficulty ··

4. $X \sim \mathrm{Uniform}(2, 8)$. Find its density, $P(3 < X < 5)$, $E[X]$ and $\mathrm{Var}(X)$. Then for $W \sim \mathrm{Uniform}(0, 0.5)$, find the density and $P(0.3 < W < 0.31)$. difficulty ·

5. Requests arrive at a server at a rate of 0.5 per minute, so the wait X for the next one is Exponential(0.5). Find the mean wait, $P(X > 4)$, the median wait and $P(X > 6 \mid X > 2)$. difficulty ··

6. Show by integration that an Exponential(λ) variable has $E[X] = \frac{1}{\lambda}$ and $\mathrm{Var}(X) = \frac{1}{\lambda^2}$. difficulty ··

7. Test scores are Normal with mean 100 and standard deviation 15. Find the fraction above 130, the fraction between 85 and 115, and the score at the 90th percentile. difficulty ··

8. $X \sim \mathrm{Uniform}(0, 1)$ and $Y = X^2$. Find the CDF and density of Y, $P(Y < 0.25)$ and $E[Y]$ two ways. difficulty ··

9. $U \sim \text{Uniform}(0,1)$. Show that $Y = -\dfrac{\ln U}{\lambda}$ is Exponential(λ), and find the sample it gives for $U = 0.3$ and $\lambda = 2$. Then find the transformation of U that samples the variable of Problem 2, and its value at $U = 0.75$. difficulty ···

10. Three independent replicas answer a request after Exponential times with rates 1, 2 and 3 per second, and the client takes the first answer. Find the distribution of the response time T, its mean, $P(T > 0.5)$, and the probability that the slowest replica (rate 1) answers first. difficulty ···

Answers

1. $c = \frac{1}{9}$, $F(x) = \frac{x^3}{27}$ on $[0, 3]$, $P(1 < X < 2) = \frac{7}{27} \approx 0.2593$
2. $f(x) = \frac{2}{x^3}$, $P(2 < X < 4) = 0.1875$, median $\sqrt{2} \approx 1.414$, $E[X] = 2$, $E[X^2] = \infty$
3. $E[X] = \frac{2}{3}$, $E[X^2] = \frac{1}{2}$, $\mathrm{Var}(X) = \frac{1}{18} \approx 0.0556$, standard deviation ≈ 0.2357, median $\frac{1}{\sqrt{2}} \approx 0.7071$
4. $f = \frac{1}{6}$, $P(3 < X < 5) = \frac{1}{3}$, $E[X] = 5$, $\mathrm{Var}(X) = 3$; $f_W = 2$, $P(0.3 < W < 0.31) = 0.02$
5. Mean 2 minutes, $P(X > 4) \approx 0.1353$, median ≈ 1.386 minutes, $P(X > 6 \mid X > 2) \approx 0.1353$
6. $E[X] = \frac{1}{\lambda}$ and $\mathrm{Var}(X) = \frac{1}{\lambda^2}$
7. $P(X > 130) \approx 0.0228$, $P(85 < X < 115) \approx 0.6827$, 90th percentile ≈ 119.2
8. $F_Y(y) = \sqrt{y}$, $f_Y(y) = \frac{1}{2\sqrt{y}}$ on $(0, 1)$, $P(Y < 0.25) = 0.5$, $E[Y] = \frac{1}{3}$
9. $Y = -\frac{\ln U}{\lambda}$ is Exponential(λ), with $y \approx 0.602$ for $U = 0.3$; for Problem 2, $X = \frac{1}{\sqrt{1-U}}$, which is 2 at $U = 0.75$
10. $T \sim \text{Exponential}(6)$, $E[T] = \frac{1}{6} \approx 0.167$ s, $P(T > 0.5) \approx 0.0498$, $P(\text{replica 1 first}) = \frac{1}{6}$

Worked solutions

Problem 1

> X has density $f(x) = cx^2$ for $0 \le x \le 3$ and 0 elsewhere. Find c, the CDF, and $P(1 < X < 2)$.

$\int_0^3 cx^2\,dx = c \cdot \dfrac{27}{3} = 9c = 1$, so $c = \frac{1}{9}$	The density must integrate to 1.
$F(x) = \int_0^x \dfrac{t^2}{9}\,dt = \dfrac{x^3}{27}$ for $0 \le x \le 3$	$F = 0$ below 0 and $F = 1$ above 3; $F(3) = 1$ is a check on c.
$P(1 < X < 2) = \int_1^2 \dfrac{x^2}{9}\,dx = \dfrac{8 - 1}{27}$	From the density.
$F(2) - F(1) = \dfrac{8}{27} - \dfrac{1}{27} = \dfrac{7}{27}$	The same number from the CDF, with no integral left to do.
$c = \frac{1}{9}$, $F(x) = \frac{x^3}{27}$ **on** $[0, 3]$, $P(1 < X < 2) = \frac{7}{27} \approx 0.2593$	The middle third of the range holds about a quarter of the probability, because the density grows like x^2.

Problem 2

> X has CDF $F(x) = 1 - \dfrac{1}{x^2}$ for $x \ge 1$ and $F(x) = 0$ for $x < 1$. Find its density, $P(2 < X < 4)$, the median, $E[X]$ and $E[X^2]$.

$f(x) = F'(x) = \dfrac{2}{x^3}$ for $x \ge 1$	Differentiate the CDF.
$P(2 < X < 4) = F(4) - F(2) = \left(1 - \dfrac{1}{16}\right) - \left(1 - \dfrac{1}{4}\right) = \dfrac{3}{16} = 0.1875$	Differences of the CDF.
$1 - \dfrac{1}{m^2} = \dfrac{1}{2}$, so $m^2 = 2$ and $m = \sqrt{2} \approx 1.4142$	The median solves $F(m) = \frac{1}{2}$.
$E[X] = \int_1^\infty x \cdot \dfrac{2}{x^3}\,dx = \int_1^\infty \dfrac{2}{x^2}\,dx = 2$	An improper integral that converges.

$f(x) = \frac{2}{x^3}$, $P(2 < X < 4) = 0.1875$, **median** $\sqrt{2} \approx 1.414$, $E[X] = 2$, $E[X^2] = \infty$	$E[X^2] = \int_1^\infty \frac{2}{x}\,dx$ diverges, so this heavy-tailed (Pareto) variable has a mean but no finite variance. The mean is pulled above the median by the long right tail.

Problem 3

> X has density $f(x) = 2x$ on $[0, 1]$. Find $E[X]$, $E[X^2]$, $\mathrm{Var}(X)$, the standard deviation and the median.

$\int_0^1 2x\,dx = 1$	Check it is a density.
$E[X] = \int_0^1 x \cdot 2x\,dx = \dfrac{2}{3}$	$\int_0^1 2x^2\,dx = \frac{2}{3}$.
$E[X^2] = \int_0^1 x^2 \cdot 2x\,dx = \dfrac{2}{4} = \dfrac{1}{2}$	Square the value, keep the density as the weight.
$\mathrm{Var}(X) = \dfrac{1}{2} - \dfrac{4}{9} = \dfrac{1}{18}$	The same shortcut as in the discrete case.
$E[X] = \frac{2}{3}$, $E[X^2] = \frac{1}{2}$, $\mathrm{Var}(X) = \frac{1}{18} \approx 0.0556$, **standard deviation** ≈ 0.2357, **median** $\frac{1}{\sqrt{2}} \approx 0.7071$	$F(x) = x^2$, so $F(m) = \frac{1}{2}$ gives $m = \frac{1}{\sqrt{2}}$. The density leans right, so the median is above the mean: the long tail is on the left.

Problem 4

> $X \sim \mathrm{Uniform}(2, 8)$. Find its density, $P(3 < X < 5)$, $E[X]$ and $\mathrm{Var}(X)$. Then for $W \sim \mathrm{Uniform}(0, 0.5)$, find the density and $P(0.3 < W < 0.31)$.

$f(x) = \dfrac{1}{8 - 2} = \dfrac{1}{6}$ on $(2, 8)$	Constant height, area 1.
$P(3 < X < 5) = \dfrac{5 - 3}{6} = \dfrac{1}{3}$	For a uniform, probability is length over total length.

$E[X] = \dfrac{2+8}{2} = 5$ and $\mathrm{Var}(X) = \dfrac{(8-2)^2}{12} = 3$	From $E[X^2] = \int_2^8 \frac{x^2}{6}\,dx = \frac{512-8}{18} = 28$ and $28 - 25 = 3$.
$f_W(w) = \dfrac{1}{0.5} = 2$ on $(0, 0.5)$	A density of 2 is fine: the interval is only 0.5 long, so the area is still 1.
$f = \frac{1}{6}$, $P(3 < X < 5) = \frac{1}{3}$, $E[X] = 5$, $\mathrm{Var}(X) = 3$; $f_W = 2$, $P(0.3 < W < 0.31) = 0.02$	$2 \times 0.01 = 0.02$: the density times a short width approximates the probability of a short interval (Mistake 1).

Problem 5

> Requests arrive at a server at a rate of 0.5 per minute, so the wait X for the next one is Exponential(0.5). Find the mean wait, $P(X > 4)$, the median wait and $P(X > 6 \mid X > 2)$.

$E[X] = \dfrac{1}{0.5} = 2$ minutes	Half an arrival per minute is one every two minutes on average.
$P(X > 4) = e^{-0.5 \cdot 4} = e^{-2} \approx 0.1353$	The survival function $e^{-\lambda x}$.
$1 - e^{-0.5m} = \dfrac{1}{2}$, so $m = \dfrac{\ln 2}{0.5} = 2\ln 2 \approx 1.3863$ minutes	The median solves $F(m) = \frac{1}{2}$.
$P(X > 6 \mid X > 2) = \dfrac{e^{-3}}{e^{-1}} = e^{-2} \approx 0.1353$	$X > 6$ is contained in $X > 2$, so the intersection is $X > 6$.
Mean 2 minutes, $P(X > 4) \approx 0.1353$, **median** ≈ 1.386 **minutes,** $P(X > 6 \mid X > 2) \approx 0.1353$	Memorylessness: after two minutes with nothing, the chance of waiting four more is the same as it was at the start. The continuous version of the geometric distribution.

Problem 6

Show by integration that an Exponential(λ) variable has $E[X] = \frac{1}{\lambda}$ and $\mathrm{Var}(X) = \frac{1}{\lambda^2}$.

$E[X] = \int_0^\infty x\lambda e^{-\lambda x}\,dx = \Big[-xe^{-\lambda x}\Big]_0^\infty + \int_0^\infty e^{-\lambda x}\,dx$	Parts with $u = x$ and $dv = \lambda e^{-\lambda x}\,dx$, so $v = -e^{-\lambda x}$.
$E[X] = 0 + \dfrac{1}{\lambda}$	$xe^{-\lambda x} \to 0$ as $x \to \infty$, and $\int_0^\infty e^{-\lambda x}\,dx = \frac{1}{\lambda}$.
$E[X^2] = \int_0^\infty x^2\lambda e^{-\lambda x}\,dx = 0 + \int_0^\infty 2xe^{-\lambda x}\,dx = \dfrac{2}{\lambda}E[X] = \dfrac{2}{\lambda^2}$	Parts again with $u = x^2$; the remaining integral is $\frac{2}{\lambda}$ times the one in step 1.
$\mathrm{Var}(X) = \dfrac{2}{\lambda^2} - \dfrac{1}{\lambda^2} = \dfrac{1}{\lambda^2}$	The shortcut.
$E[X] = \frac{1}{\lambda}$ **and** $\mathrm{Var}(X) = \frac{1}{\lambda^2}$	The standard deviation equals the mean. For $\lambda = 0.5$ in Problem 5, both are 2 minutes.

Problem 7

Test scores are Normal with mean 100 and standard deviation 15. Find the fraction above 130, the fraction between 85 and 115, and the score at the 90th percentile.

$z = \dfrac{130 - 100}{15} = 2$	Standardize: subtract the mean, divide by the standard deviation.
$P(X > 130) = 1 - \Phi(2) = 1 - 0.97725 \approx 0.0228$	About 2.3%.
$P(85 < X < 115) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \approx 0.6827$	$z = \pm 1$; by symmetry $\Phi(-1) = 1 - \Phi(1)$.
$x = 100 + 15 \times 1.2816 \approx 119.22$	Solve $\Phi(z) = 0.9$ for $z = 1.2816$, then undo the standardization: $x = \mu + \sigma z$.

$P(X > 130) \approx 0.0228$, $P(85 < X < 115) \approx 0.6827$, **90th percentile** ≈ 119.2	Within one standard deviation of the mean lies about 68% of a normal distribution, and beyond two standard deviations on one side lies about 2.3%.

Problem 8

> $X \sim \text{Uniform}(0, 1)$ and $Y = X^2$. Find the CDF and density of Y, $P(Y < 0.25)$ and $E[Y]$ two ways.

$F_Y(y) = P(X^2 \le y) = P(X \le \sqrt{y}) = \sqrt{y}$ for $0 \le y \le 1$	$X \ge 0$, so $X^2 \le y$ exactly when $X \le \sqrt{y}$.
$f_Y(y) = \dfrac{d}{dy}\sqrt{y} = \dfrac{1}{2\sqrt{y}}$ on $(0, 1)$	Equivalently $f_X(\sqrt{y})\left\lvert\frac{dx}{dy}\right\rvert = 1 \cdot \frac{1}{2\sqrt{y}}$. It is unbounded near 0 and still integrates to 1.
$P(Y < 0.25) = F_Y(0.25) = 0.5$	$Y < 0.25$ exactly when $X < 0.5$.
$E[Y] = \int_0^1 y \cdot \dfrac{1}{2\sqrt{y}}\,dy = \int_0^1 \dfrac{\sqrt{y}}{2}\,dy = \dfrac{1}{3}$	From the density of Y.
$F_Y(y) = \sqrt{y}$, $f_Y(y) = \frac{1}{2\sqrt{y}}$ **on** $(0, 1)$, $P(Y < 0.25) = 0.5$, $E[Y] = \frac{1}{3}$	Check without f_Y: $E[X^2] = \int_0^1 x^2\,dx = \frac{1}{3}$. Squaring numbers in $(0, 1)$ crowds them toward 0, which is where the density piles up.

Problem 9

> $U \sim \text{Uniform}(0, 1)$. Show that $Y = -\dfrac{\ln U}{\lambda}$ is Exponential(λ), and find the sample it gives for $U = 0.3$ and $\lambda = 2$. Then find the transformation of U that samples the variable of Problem 2, and its value at $U = 0.75$.

$P(Y > y) = P\big(-\ln U > \lambda y\big) = P\big(U < e^{-\lambda y}\big)$	Multiply by -1 and exponentiate; e^x is increasing, so the inequality keeps its direction.
$P(Y > y) = e^{-\lambda y}$ for $y \ge 0$	For a uniform, $P(U < u) = u$. That is the exponential survival function.

$U = 0.3$, $\lambda = 2$: $y = -\dfrac{\ln 0.3}{2} \approx 0.6020$	One uniform random number becomes one exponential sample.
In general $X = F^{-1}(U)$ has CDF F: $P(F^{-1}(U) \le x) = P(U \le F(x)) = F(x)$	Inverse-CDF sampling. For the exponential, $F^{-1}(u) = -\frac{\ln(1-u)}{\lambda}$, and $1 - U$ is uniform too, which gives the formula above.
$Y = -\frac{\ln U}{\lambda}$ **is Exponential**(λ)**, with** $y \approx 0.602$ **for** $U = 0.3$**; for Problem 2,** $X = \frac{1}{\sqrt{1-U}}$**, which is** 2 **at** $U = 0.75$	Solve $u = 1 - \frac{1}{x^2}$ for x. This is how libraries turn a uniform random number generator into samples from other distributions.

Problem 10

> Three independent replicas answer a request after Exponential times with rates 1, 2 and 3 per second, and the client takes the first answer. Find the distribution of the response time T, its mean, $P(T > 0.5)$, and the probability that the slowest replica (rate 1) answers first.

$P(T > t) = P(X_1 > t)P(X_2 > t)P(X_3 > t) = e^{-t}e^{-2t}e^{-3t} = e^{-6t}$	The minimum is above t only if all three are; independence lets the probabilities multiply.
$T \sim \text{Exponential}(6)$, so $E[T] = \dfrac{1}{6}$ second	The rates add.
$P(T > 0.5) = e^{-3} \approx 0.0498$	The survival function at 0.5.
$P(X_1 = T) = \int_0^\infty e^{-t} \cdot e^{-5t}\,dt = \dfrac{1}{6}$	Replica 1 answers at t (density e^{-t}) and the other two are still busy (probability $e^{-2t}e^{-3t}$).
$T \sim \text{Exponential}(6)$, $E[T] = \frac{1}{6} \approx 0.167$ **s,** $P(T > 0.5) \approx 0.0498$, $P(\text{replica 1 first}) = \frac{1}{6}$	In general each replica wins with probability its rate over the total, $\frac{\lambda_i}{\lambda_1+\lambda_2+\lambda_3}$, so the others win with $\frac{2}{6}$ and $\frac{3}{6}$.

Where this goes wrong

1. The density read as a probability

In the discrete case $p(x) = P(X = x)$, and the density looks like it should mean the same.

$W \sim \text{Uniform}(0, 0.5)$ has density $f_W(w) = 2$	Right so far: Problem 4.
"The density at 0.3 is the probability of 0.3."	The step that causes the mistake: a density is probability per unit length, not probability.
$P(W = 0.3) = f_W(0.3) = 2$ ✗	A probability can never be 2. For a continuous variable $P(W = 0.3) = 0$; the density only becomes a probability when multiplied by a width, so $P(0.3 < W < 0.31) \approx 2 \times 0.01 = 0.02$.

2. Change of variables without |dx/dy|

The density of $Y = g(X)$ is found by substituting the inverse into the density of X.

$X \sim \text{Uniform}(0, 1)$, $Y = X^2$, and $x = \sqrt{y}$	Right so far: Problem 8.
"$f_Y(y) = f_X(\sqrt{y}) = 1$."	The step that causes the mistake: it moves the height of the density but not the width of the intervals, which squaring shrinks near 0.
Y is Uniform$(0, 1)$, so $P(Y < 0.25) = 0.25$ ✗	But $Y < 0.25$ is exactly $X < 0.5$, which has probability 0.5. The factor $\left\lvert \frac{dx}{dy} \right\rvert = \frac{1}{2\sqrt{y}}$ is what keeps probabilities equal, and going through the CDF, $F_Y(y) = \sqrt{y}$, never loses it.

3. The exponential's mean taken as λ

The rate is the parameter in the formula, and it is read as the average.

Requests arrive at $\lambda = 0.5$ per minute, and the wait is Exponential(0.5)	Right so far: Problem 5.

“The parameter of the distribution is its mean.”	True of the Poisson count, not of the exponential wait.
$E[X] = \lambda = 0.5$ minutes ✗	Half an arrival per minute means a wait of $\frac{1}{0.5} = 2$ minutes on average: $E[X] = \frac{1}{\lambda}$, as Problem 6 shows. A check by units: λ is per minute, and a wait has to come out in minutes.

4. Standardizing with the variance

The normal is written Normal(μ, σ^2), and the second parameter goes in the denominator.

Scores are Normal$(100, 15^2)$: $\mu = 100$, $\sigma^2 = 225$	Right so far: Problem 7.
“$z = \dfrac{x - \mu}{\sigma^2}$.”	The step that causes the mistake: Z has to be unitless, and $x - \mu$ is in points while σ^2 is in points squared.
$z = \frac{30}{225} \approx 0.1333$, so $P(X > 130) = 1 - \Phi(0.1333) \approx 0.4470$ ✗	Nearly half the class above 130 would put 130 barely above average. Divide by the standard deviation: $z = \frac{30}{15} = 2$ and $P(X > 130) \approx 0.0228$.

5. The mean taken as the median

For a normal, half the probability is below the mean, and the same is assumed of every distribution.

The wait in Problem 5 has mean 2 minutes	Right so far.
“Half the waits are shorter than the average wait.”	True for a symmetric distribution; the exponential is skewed right.

$P(X < 2) = 0.5$ ✗	$P(X < 2) = 1 - e^{-1} \approx 0.6321$: the long right tail pulls the mean up, so most waits are shorter than average. The median is $2 \ln 2 \approx 1.386$ minutes.

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 5 of *Probability & Statistics for Machine Learning*, a 250-problem workbook with worked solutions: eulerpad.app/workbooks

eulerpad.app/practice/continuous-random-variables-and-densities