

CSE 312 – Section 6 Solutions

Summer 2026

Review of Main Concepts

- **Continuous Random Variable:** A continuous random variable X has an uncountably infinite number of values and its cumulative distribution function $F_X(x) : \mathbb{R} \rightarrow \mathbb{R}$ is continuous everywhere.
- **Cumulative Distribution Function (cdf):** For any random variable (discrete or continuous) X , the cumulative distribution function is defined as $F_X(x) = \mathbb{P}(X \leq x)$. Notice that (1) this function must be monotonically nondecreasing: if $x < y$ then $F_X(x) \leq F_X(y)$, because $\mathbb{P}(X \leq x) \leq \mathbb{P}(X \leq y)$; (2) since probabilities are between 0 and 1, that $0 \leq F_X(x) \leq 1$ for all x , with $\lim_{x \rightarrow -\infty} F_X(x) = 0$ and $\lim_{x \rightarrow +\infty} F_X(x) = 1$; (3) since $\mathbb{P}(X = k) = 0$ for some constant k if X is a continuous random variable, $\mathbb{P}(X < k) = \mathbb{P}(X \leq k)$.
- **Probability Density Function (pdf or density):** Let X be a continuous random variable. Then the probability density function $f_X(x) : \mathbb{R} \rightarrow \mathbb{R}$ of X is defined as $f_X(x) = \frac{d}{dx} F_X(x)$. Taking the integral of both sides, it means that $F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(t) dt$. It follows that $\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x) dx$ and that $\int_{-\infty}^{\infty} f_X(x) dx = 1$. From the fact that $F_X(x)$ is monotonically nondecreasing it follows that $f_X(x) \geq 0$ for every real number x . Note that $f_X(a) \neq \mathbb{P}(X = a)$, since $\mathbb{P}(X = a) = F_X(a) - F_X(a) = 0$ for all a . However, the probability that X is close to a is proportional to $f_X(a)$: for small δ , $\mathbb{P}(a - \frac{\delta}{2} < X < a + \frac{\delta}{2}) \approx \delta f_X(a)$.
- **i.i.d. (independent and identically distributed):** Random variables $X_1, \dots, X_n$ are i.i.d. (or iid) if they are independent and have the same probability mass function or probability density function.
- **Discrete to Continuous:** To summarize, when going from discrete to continuous, the main differences are (1) using an integral instead of a summation, and (2) using the density function $f_X(k)$ instead of the PMF $\mathbb{P}(X = k)$.

	Discrete	Continuous
PMF/PDF	$p_X(x) = \mathbb{P}(X = x)$	$f_X(x) \neq \mathbb{P}(X = x) = 0$
CDF	$F_X(x) = \sum_{t \leq x} p_X(t)$	$F_X(x) = \int_{-\infty}^x f_X(t) dt$
Normalization	$\sum_x p_X(x) = 1$	$\int_{-\infty}^{\infty} f_X(x) dx = 1$
Expectation	$\mathbb{E}[X] = \sum_x x p_X(x)$	$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
LOTUS	$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x)$	$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$

- **Standardizing:** Let X be any random variable (discrete or continuous, not necessarily normal), with $\mathbb{E}[X] = \mu$ and $\text{Var}(X) = \sigma^2$. If we let $Y = \frac{X - \mu}{\sigma}$, then $\mathbb{E}[Y] = 0$ and $\text{Var}(Y) = 1$.

- **Law of Total Probability (Continuous):** This may not have been covered in class yet, but will be at some point, and you will use it on the problem set. A is an event, and X is a continuous random variable with density function $f_X(x)$.

$$\mathbb{P}(A) = \int_{-\infty}^{\infty} \mathbb{P}(A \mid X = x) f_X(x) dx$$

- **Transforming Continuous Random Variables** (May not be covered in class.) Suppose that X is a discrete random variable that takes values in Ω_X and let $Y = g(X)$ for some function g . Let $\Omega_Y = \{g(x) \mid x \in \Omega_X\}$. Then the probability mass function of Y satisfies

$$p_Y(y) = \sum_{x \in \Omega_X \mid g(x)=y} p_X(x).$$

However, if X is a continuous random variable with density function f , and $Y = g(X)$ for some continuous function g , then we can **not** say that

$$f_Y(y) = \int_{x \in \Omega_X \mid g(x)=y} f_X(x) dx.$$

Rather, we must take the following steps:

- Compute $F_Y(y)$ from $F_X(x)$.
- Differentiate $F_Y(y)$ with respect to y to obtain $f_Y(y)$.

See Problems 17 and 18.

- **Zoo of Continuous Random Variables**

- Uniform:** $X \sim \text{Uniform}(a, b)$ iff X has the following probability density function:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

$\mathbb{E}[X] = \frac{a+b}{2}$ and $\text{Var}(X) = \frac{(b-a)^2}{12}$. This represents each real number from $[a, b]$ to be equally likely.

- Exponential:** $X \sim \text{Exponential}(\lambda)$ iff X has the following probability density function:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

$\mathbb{E}[X] = \frac{1}{\lambda}$ and $\text{Var}(X) = \frac{1}{\lambda^2}$. $F_X(x) = 1 - e^{-\lambda x}$ for $x \geq 0$. The exponential random variable is the continuous analog of the geometric random variable: it represents the waiting time to the next event, where $\lambda > 0$ is the average number of events per unit time. Note that the exponential measures how much time passes until the next event (any real number, continuous), whereas the Poisson measures how many events occur in a unit of time (nonnegative integer, discrete). The exponential random variable X is memoryless:

$$\text{for any } s, t \geq 0, \mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$$

The geometric random variable also has this property.

- c) **Normal (Gaussian, “bell curve”)**: $X \sim \mathcal{N}(\mu, \sigma^2)$ iff X has the following probability density function:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}, \quad x \in \mathbb{R}$$

$\mathbb{E}[X] = \mu$ and $\text{Var}(X) = \sigma^2$. The “standard normal” random variable is typically denoted Z and has mean 0 and variance 1. The CDF has no closed form, but we denote the CDF of the standard normal as $\Phi(z) = F_Z(z) = \mathbb{P}(Z \leq z)$. Note from symmetry of the probability density function about $z = 0$ that: $\Phi(-z) = 1 - \Phi(z)$.

To find the values of $\Phi(\cdot)$, you can use this [Z-table](#).

Closure of the Normal Distribution: Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then, $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$. That is, linear transformations of normal random variables are still normal. Thus, for example, if $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$.

“Reproductive” Property of Normals: Let $X_1, \dots, X_n$ be independent normal random variables with $\mathbb{E}[X_i] = \mu_i$ and $\text{Var}(X_i) = \sigma_i^2$. Let $a_1, \dots, a_n \in \mathbb{R}$ and $b \in \mathbb{R}$. Then,

$$X = \sum_{i=1}^n (a_i X_i + b) \sim \mathcal{N}\left(\sum_{i=1}^n (a_i \mu_i + b), \sum_{i=1}^n a_i^2 \sigma_i^2\right)$$

There’s nothing special about the parameters – the important result here is that the resulting random variable is still normally distributed.

- **Central Limit Theorem (CLT):** Let $X_1, \dots, X_n$ be iid random variables with $\mathbb{E}[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$. Let $X = \sum_{i=1}^n X_i$, which has $\mathbb{E}[X] = n\mu$ and $\text{Var}(X) = n\sigma^2$. Let $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, which has $\mathbb{E}[\bar{X}] = \mu$ and $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$. $\bar{X}$ is called the *sample mean*. Then, as $n \rightarrow \infty$, $\bar{X}$ approaches the normal distribution $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. Standardizing, this is equivalent to $Y = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ approaching $\mathcal{N}(0, 1)$. Similarly, as $n \rightarrow \infty$, X approaches $\mathcal{N}(n\mu, n\sigma^2)$ and $Y' = \frac{X-n\mu}{\sigma\sqrt{n}}$ approaches $\mathcal{N}(0, 1)$. It is no surprise that $\bar{X}$ has mean μ and variance σ^2/n – we have seen this before and it is easy to show. The importance of the CLT is that, for large n , regardless of what distribution X_i comes from, $\bar{X}$ is *approximately normally distributed with mean μ and variance σ^2/n* .

Announcements & Plan for Section

Announcements

- PSet 4 was due yesterday.
- Pset 5 is released - will be next week on 8/5.
- This week’s focus: continuous distributions and CLT

Plan for Section

- Content Review (Problem 1)

- Problem 7
- Problem 9
- Problem 12
- Problem 20

1 Content Review

a) What is $\mathbb{P}(X = 4)$ if X is a **continuous** random variable?

- ☐ 1
- ☐ 0
- ☐ not enough information

(b). If X is a continuous random variable, the probability it takes on a particular constant is 0 since the support of X has infinite real values.

b) The cumulative distribution function for a continuous random variable X is $F_X(k) =$

- ☐ $\int_{-\infty}^k f_X(x)dx$
- ☐ $\int_{-\infty}^{\infty} f_X(x)dx$
- ☐ $\int_k^{\infty} f_X(x)dx$
- ☐ $\frac{d}{dk}f_X(k)$

(a) We take the integral over the PDF over the appropriate range to get the CDF. Since the CDF is $F_X(k) = \mathbb{P}(X \leq k)$ we take the integral from negative infinity up to k .

c) The probability density function for a continuous random variable X is $f_X(k) =$

- ☐ $\int_{-\infty}^k f_X(x)dx$
- ☐ $\frac{d}{dk}F_X(k)$

(b) We take the derivative of the CDF to get the PDF.

d) **True or False.** If X is a continuous random variable, $\mathbb{E}[X] = \int_{-\infty}^{\infty} xf_X(x)dx$

True. This is by definition of expectation for continuous random variables. Note the difference from the discrete case is that we're using an integral instead of a summation, and we're using density instead of pmf!

e) **True or False.** If X is a continuous random variable, $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

True. This definition for variance applies regardless of whether X is discrete or continuous.

f) Which of the following follow an $\text{Exponential}(\lambda)$ distribution?

- ☐ Number of minutes to the first success with λ as average number of successes per minute
- ☐ Number of successes in the first 1 minute with λ as average number of successes per minute
- ☐ Time (real number) to the first success with λ as average number of successes per minute

(c) The exponential random variable is from our zoo of continuous random variables, and represents the time (continuous) till the first success. Note that (b) is a Poisson random variable with parameter λ !

g) True or False: For any random variable X , $\mathbb{P}(X = 5) = \mathbb{P}(X - 5 = 0)$.

True. We can think of $X - 5$ as another random variable where we take the output of X and subtract five from it. Then the probability that $X - 5$ is zero is identical to the probability that X is originally five.

h) True or False: For some continuous random variable X , $\mathbb{P}(X \leq 5) \neq \mathbb{P}(X < 5)$.

False. Note that $\mathbb{P}(X \leq 5) = \mathbb{P}(X = 5) + \mathbb{P}(X < 5)$. But the first term is zero, so the probabilities are exactly equal. This holds for every continuous random variable.

i) True or False: Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$. Then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

True. This follows by the closure of the normal distribution.

2 Uniform2

Robbie decided he wanted to create a “new” type of distribution that will be famous, but he needs some help. He knows he wants it to be continuous and have uniform density, but he needs help working out some of the details. We’ll denote a random variable X having the “Uniform-2” distribution as $X \sim \text{Uniform2}(a, b, c, d)$, where $a < b < c < d$. We want the density to be non-zero in $[a, b]$ and $[c, d]$, and zero everywhere else. Anywhere the density is non-zero, it must be equal to the same constant.

- a) Find the probability density function, $f_X(x)$. Be sure to specify the values (in terms of a, b, c, d) it takes on for every point in $(-\infty, \infty)$. (Hint: use a piecewise definition).

We want our probability density function to have a non-zero, uniform density in the intervals $[a, b]$ and $[c, d]$, and zero everywhere else. Let ℓ be that value. Then

$$f_X(x) = \begin{cases} \ell & \text{if } a \leq x \leq b \text{ or } c \leq x \leq d \\ 0 & \text{otherwise .} \end{cases}$$

In order for this to be a valid probability density function, we must have $\int_{-\infty}^{\infty} f_X(x) dx = 1$. Solving,

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f_X(x) dx = \int_{-\infty}^a f_X(x) dx + \int_a^b f_X(x) dx + \int_b^c f_X(x) dx \\ &\quad + \int_c^d f_X(x) dx + \int_d^{\infty} f_X(x) dx \\ &= 0 + \int_a^b \ell dx + 0 + \int_c^d \ell dx + 0 \\ &= \ell(b-a) + \ell(d-c) . \end{aligned}$$

Note that taking the integral over any range that is not $[a, b]$ and $[c, d]$ gives a zero output since $f_X(x) = 0$ outside of those ranges. Rearranging, we get $\ell = \frac{1}{(b-a)+(d-c)}$, so

$$f_X(x) = \begin{cases} \frac{1}{(b-a)+(d-c)} & \text{if } a \leq x \leq b \text{ or } c \leq x \leq d \\ 0 & \text{otherwise .} \end{cases}$$

- b) Find the cumulative distribution function, $F_X(x)$. Be sure to specify the values it takes on for every point in $(-\infty, \infty)$. (Hint: use a piecewise definition).

Let's keep using ℓ from before to reduce clutter. Recall that the cumulative distribution function takes an integral over the probability density function from negative infinity up to x , i.e.,

$$F_X(x) = \int_{-\infty}^x f_X(t) dt .$$

We use t as our step variable here to not overload the variable x . Let us consider possible values that x can take on. For $x < a$,

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = 0 .$$

For $a \leq x < b$,

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \int_{-\infty}^a f_X(t) dt + \int_a^x f_X(t) dt = \ell(x-a) .$$

For $b \leq x < c$,

$$F_X(x) = \int_{-\infty}^x f_X(t) dt = \int_{-\infty}^a f_X(t) dt + \int_a^b f_X(t) dt + \int_b^x f_X(t) dt = \ell(b-a) .$$

For $c \leq x < d$,

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x f_X(t) dt = \int_{-\infty}^a f_X(t) dt + \int_a^b f_X(t) dt + \int_b^c f_X(t) dt + \int_c^x f_X(t) dt \\ &= \ell(b-a) + \ell(x-c) . \end{aligned}$$

Finally, for $x \geq d$,

$$\begin{aligned} F_X(x) &= \int_{-\infty}^x f_X(t) dt = \int_{-\infty}^a f_X(t) dt + \int_a^b f_X(t) dt + \int_b^c f_X(t) dt \\ &\quad + \int_c^d f_X(t) dt + \int_d^x f_X(t) dt \\ &= \ell(b-a) + \ell(d-c) . \end{aligned}$$

Putting everything together (and now substituting $\ell = \frac{1}{(b-a)+(d-c)}$),

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{(b-a)+(d-c)} & \text{if } a \leq x < b \\ \frac{b-a}{(b-a)+(d-c)} & \text{if } b \leq x < c \\ \frac{(b-a)+(x-c)}{(b-a)+(d-c)} & \text{if } c \leq x < d \\ 1 & \text{if } x \geq d . \end{cases}$$

3 Create the distribution

Suppose X is a continuous random variable that is uniform on $[0, 1)$ and uniform on $[1, 2]$, but

$$\mathbb{P}(1 \leq X \leq 2) = 2 \cdot \mathbb{P}(0 \leq X < 1).$$

Outside of $[0, 2]$ the density is 0. What is the PDF and CDF of X ?

The fact that X is uniform on each of the intervals means that its PDF is constant on each. So,

$$f_X(x) = \begin{cases} c & 0 \leq x < 1 \\ d & 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Taking the integral of the PDF yields the CDF. For $0 \leq x < 1$:

$$F_X(x) = \int_0^x c dt = cx$$

For $1 \leq x \leq 2$:

$$F_X(x) = \int_0^1 c dt + \int_1^x d dt = c + d(x-1) = dx + c - d$$

So the CDF is:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ cx & 0 \leq x < 1 \\ dx + c - d & 1 \leq x \leq 2 \\ 1 & x > 2 \end{cases}$$

To solve for c and d , we use the provided condition that $\mathbb{P}(1 \leq X \leq 2) = 2 \cdot \mathbb{P}(0 \leq X < 1)$, which is equivalent to:

$$F_X(2) - F_X(1) = 2 \cdot (F_X(1) - F_X(0))$$

Plugging in the expressions from our CDF gives:

$$d = 2c$$

We also know that the total probability must sum to 1, so the CDF at $x = 2$ must be 1:

$$F_X(2) = 2d + c - d = c + d = 1$$

Solving this system of equations ($d = 2c$ and $c + d = 1$) yields $c = \frac{1}{3}$ and $d = \frac{2}{3}$.

Substituting these values back into our PDF and CDF gives the final functions:

$$f_X(x) = \begin{cases} 1/3 & 0 \leq x < 1 \\ 2/3 & 1 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{3}x & 0 \leq x < 1 \\ \frac{2}{3}x - \frac{1}{3} & 1 \leq x \leq 2 \\ 1 & x > 2 \end{cases}$$

4 The Spotlight

A spotlight is mounted on a wall at a height h above the ground. The light rotates and is equally likely to point at any angle Θ between 0 and $\pi/4$ (where 0 corresponds to pointing straight down at the ground). Let X be the distance along the ground from the point directly beneath the light to the spot where the light hits the ground. (Note that $X = h \tan(\Theta)$). For X , find ...

- a) the cumulative distribution function F_X .

Hint: First, determine the range of possible values for X given the bounds on Θ . Then, use the definition $F_X(x) = \mathbb{P}[X \leq x]$ and substitute for X . Recall that if $\tan(x) = y$, then $x = \arctan(y)$.

We are given that $\Theta \sim \text{Unif}(0, \pi/4)$. Therefore, the CDF of Θ is $F_\Theta(\theta) = \frac{\theta}{\pi/4} = \frac{4\theta}{\pi}$ for $0 \leq \theta \leq \pi/4$. Since $X = h \tan(\Theta)$, the range of possible values for X is

$[h \tan(0), h \tan(\pi/4)] = [0, h]$. For $x \in [0, h]$, we find the CDF of X as follows:

$$F_X(x) = \mathbb{P}[X \leq x] = \mathbb{P}[h \tan(\Theta) \leq x] = \mathbb{P}[\Theta \leq \arctan(x/h)]$$

Substituting the CDF of Θ :

$$F_X(x) = \frac{4}{\pi} \arctan\left(\frac{x}{h}\right)$$

Thus, our CDF is:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{4}{\pi} \arctan\left(\frac{x}{h}\right) & 0 \leq x \leq h \\ 1 & x > h \end{cases}$$

b) the probability density function f_X .

Hint: Recall the chain rule for derivatives, and that $\frac{d}{du} \arctan(u) = \frac{1}{1+u^2}$.

We know that $f_X(x) = \frac{d}{dx} F_X(x)$ for $x \in [0, h]$. Using the chain rule:

$$f_X(x) = \frac{d}{dx} \left[\frac{4}{\pi} \arctan\left(\frac{x}{h}\right) \right] = \frac{4}{\pi} \cdot \frac{1}{1 + (x/h)^2} \cdot \frac{1}{h} = \frac{4h}{\pi(h^2 + x^2)}$$

Thus, our PDF is:

$$f_X(x) = \begin{cases} \frac{4h}{\pi(h^2 + x^2)} & 0 \leq x \leq h \\ 0 & \text{otherwise} \end{cases}$$

c) the expected value $\mathbb{E}[X]$.

Hint: To evaluate the integral $\int \frac{x}{h^2 + x^2} dx$, try using the substitution $u = h^2 + x^2$.

We compute the expected value using the PDF:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \int_0^h x \cdot \frac{4h}{\pi(h^2 + x^2)} dx$$

Using the substitution $u = h^2 + x^2$, we have $du = 2x dx$:

$$\mathbb{E}[X] = \frac{2h}{\pi} \int_{h^2}^{2h^2} \frac{1}{u} du = \frac{2h}{\pi} \ln(u) \Big|_{h^2}^{2h^2} = \frac{2h}{\pi} (\ln(2h^2) - \ln(h^2)) = \frac{2h \ln(2)}{\pi}.$$

d) the variance $\text{Var}(X)$.

Hint: To find $\mathbb{E}[X^2]$, you will need to integrate a fraction like $\frac{x^2}{h^2 + x^2}$. Try adding and subtracting h^2 in the numerator to split the fraction into two easier pieces: $x^2 = x^2 + h^2 - h^2$.

First, we compute the second moment. Using the hint, we can simplify the integrand:

$$\mathbb{E}[X^2] = \int_0^h x^2 \cdot \frac{4h}{\pi(h^2 + x^2)} dx = \frac{4h}{\pi} \int_0^h \frac{x^2 + h^2 - h^2}{x^2 + h^2} dx$$

$$\mathbb{E}[X^2] = \frac{4h}{\pi} \left(\int_0^h 1 dx - \int_0^h \frac{h^2}{x^2 + h^2} dx \right)$$

The second integral is exactly the form of the derivative we found in part 2, which integrates back to our arctangent function:

$$\mathbb{E}[X^2] = \frac{4h}{\pi} \left(h - h \arctan\left(\frac{x}{h}\right) \Big|_0^h \right) = \frac{4h}{\pi} \left(h - h \left(\frac{\pi}{4}\right) \right) = \frac{4h^2}{\pi} - h^2.$$

Consequently, the variance is:

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \left(\frac{4h^2}{\pi} - h^2 \right) - \left(\frac{2h \ln(2)}{\pi} \right)^2 = h^2 \left(\frac{4}{\pi} - 1 - \frac{4(\ln 2)^2}{\pi^2} \right).$$

5 Max of uniforms

Let $U_1, U_2, \dots, U_n$ be mutually independent Uniform random variables on $(0, 1)$. As in the discrete case, independence of these random variables implies that

$$\mathbb{P}(U_1 \leq x_1, \dots, U_n \leq x_n) = \mathbb{P}(U_1 \leq x_1) \cdots \mathbb{P}(U_n \leq x_n)$$

for any numbers $x_1, \dots, x_n$. Find the CDF and PDF for the random variable $Z = \max(U_1, \dots, U_n)$.

The key observation is that the max of n numbers $\max(a_1, \dots, a_n)$ is less than or equal to some constant x , if and only if each individual number is less than or equal to that constant x (i.e. $a_i \leq x$ for all i). Using this idea, we get:

$$\begin{aligned} F_Z(x) &= \mathbb{P}(Z \leq x) = \mathbb{P}(\max(U_1, \dots, U_n) \leq x) \\ &= \mathbb{P}(U_1 \leq x, \dots, U_n \leq x) \\ &= \mathbb{P}(U_1 \leq x) \cdots \mathbb{P}(U_n \leq x) && \text{[independence]} \\ &= F_{U_1}(x) \cdots F_{U_n}(x) \\ &= F_U(x)^n && \text{[where } U \sim \text{Uniform}(0, 1)\text{]} \end{aligned}$$

So the CDF of Z is

$$F_Z(x) = \begin{cases} 0 & x < 0 \\ x^n & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

To find the PDF, we take the derivative of each part of the CDF, which gives us the

following:

$$f_Z(x) = \begin{cases} nx^{n-1} & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

6 New PDF?

Alex came up with a function that he thinks could represent a probability density function. He defined the potential PDF for X as $f(x) = \frac{1}{1+x^2}$ defined on $[0, \infty)$. Is this a valid PDF? If not, find a constant c such that the PDF $f_X(x) = \frac{c}{1+x^2}$ is valid. Then find $\mathbb{E}[X]$. (Hints: $\frac{d}{dx}(\tan^{-1} x) = \frac{1}{1+x^2}$, $\tan \frac{\pi}{2} = \infty$, and $\tan 0 = 0$.)

$f(x) = \frac{1}{1+x^2}$ is not a valid PDF:

$$\int_0^\infty \frac{1}{1+x^2} dx = \tan^{-1} x \Big|_0^\infty = \left(\frac{\pi}{2} - 0 \right) = \frac{\pi}{2} \neq 1$$

. The area under the PDF must be 1. So,

$$\int_0^\infty \frac{c}{1+x^2} dx = c \tan^{-1} x \Big|_0^\infty = c \left(\frac{\pi}{2} - 0 \right) = 1$$

Solving for c gives us $c = 2/\pi$.

Using our value we found for c , and the definition of expectation we can compute $\mathbb{E}[X]$ as follows:

$$\mathbb{E}[X] = \int_0^\infty \frac{cx}{1+x^2} dx = \frac{2}{\pi} \int_0^\infty \frac{x}{1+x^2} dx = \frac{1}{\pi} \ln(1+x^2) \Big|_0^\infty = \infty$$

7 Throwing a dart

Consider the closed unit circle of radius r , i.e., $S = \{(x, y) : x^2 + y^2 \leq r^2\}$. Suppose we throw a dart onto this circle and are guaranteed to hit it, but the dart is equally likely to land anywhere in S . Concretely this means that the probability that the dart lands in any particular area of size A (that is entirely inside the circle of radius R), is equal to $\frac{A}{\text{Area of whole circle}}$. The density outside the circle of radius r is 0.

Let X be the distance the dart lands from the center. What is the CDF and pdf of X ? What is $\mathbb{E}[X]$ and $\text{Var}(X)$?

Since $F_X(x)$ is the probability that the dart lands inside the circle of radius x , that probability is the area of a circle of radius x divided by the area of the circle of radius r (i.e., $\pi x^2 / \pi r^2$).

Thus, our CDF looks like

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{r^2} & 0 \leq x \leq r \\ 1 & x > r \end{cases}$$

To find the PDF we just need to take the derivative of the CDF, which gives us the following:

$$f_X(x) = \begin{cases} \frac{2x}{r^2} & 0 < x \leq r \\ 0 & \text{otherwise} \end{cases}$$

Using the definition of expectation we get

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^r x \frac{2x}{r^2} dx = \frac{2}{3r^2} [x^3]_0^r = \frac{2}{3}r$$

We know that $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^r x^2 \frac{2x}{r^2} dx = \frac{2}{4r^2} [x^4]_0^r = \frac{1}{2}r^2$$

Plugging this into our variance equation gives

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{2}r^2 - \left(\frac{2}{3}r\right)^2 = \frac{1}{18}r^2$$

8 A square dartboard?

You throw a dart at an $s \times s$ square dartboard. The goal of this game is to get the dart to land as close to the lower left corner of the dartboard as possible. However, your aim is such that the dart is equally likely to land at any point on the dartboard. Let random variable X be the length of the side of the smallest *square* B in the lower left corner of the dartboard that contains the point where the dart lands. That is, the lower left corner of B must be the same point as the lower left corner of the dartboard, and the dart lands somewhere along the upper or right edge of B . For X , find the CDF, PDF, $\mathbb{E}[X]$, and $\text{Var}(X)$.

Since $F_X(x)$ is the probability that the dart lands inside the square of side length x , that probability is the area of a square of length x divided by the area of the square dartboard of length s (i.e., x^2/s^2). Thus, our CDF looks like

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ x^2/s^2, & \text{if } 0 \leq x \leq s \\ 1, & \text{if } x > s \end{cases}$$

To find the PDF, we just need to take the derivative of the CDF, which gives us the following:

$$f_X(x) = \frac{d}{dx} F_X(x) = \begin{cases} 2x/s^2, & \text{if } 0 \leq x \leq s \\ 0, & \text{otherwise} \end{cases}$$

Using the definition of expectation and variance we can compute $\mathbb{E}[X]$ and $\text{Var}(X)$ in the following manner:

$$\begin{aligned}\mathbb{E}[X] &= \int_0^s x f_X(x) dx = \int_0^s \frac{2x^2}{s^2} dx = \frac{2}{s^2} \int_0^s x^2 dx = \frac{2}{3s^2} [x^3]_0^s = \frac{2}{3}s \\ \mathbb{E}[X^2] &= \int_0^s x^2 f_X(x) dx = \int_0^s \frac{2x^3}{s^2} dx = \frac{2}{s^2} \int_0^s x^3 dx = \frac{1}{2s^2} [x^4]_0^s = \frac{1}{2}s^2 \\ \text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{2}s^2 - \left(\frac{2}{3}s\right)^2 = \frac{1}{18}s^2\end{aligned}$$

9 Will the battery last?

Suppose that the number of miles that a car can run before its battery wears out is exponentially distributed with expectation 10,000 miles. If the owner wants to take a 5000 mile road trip, what is the probability that she will be able to complete the trip without replacing the battery, given that the car has already been used for 2000 miles on the road trip?

Let N be a random variable denoting the number of miles until the battery wears out. Then $N \sim \text{Exponential}(10,000^{-1})$, because N measures the “time” (in this case miles) before an occurrence (the battery wears out) and has expectation 10,000. Since this is an exponential distribution, and the expectation of an exponential distribution is $\frac{1}{\lambda}$, $\lambda = \frac{1}{10,000}$. Therefore, via the property of memorylessness of the exponential distribution:

$$\mathbb{P}(N \geq 5000 \mid N \geq 2000) = \mathbb{P}(N \geq 3000) = 1 - \mathbb{P}(N \leq 3000) = 1 - \left(1 - e^{-\frac{3000}{10000}}\right) \approx 0.741$$

10 Batteries and exponential distributions

Let X_1, X_2 be independent exponential random variables, where X_i has parameter λ_i , for $1 \leq i \leq 2$. Let $Y = \min(X_1, X_2)$.

- a) Show that Y is an exponential random variable with parameter $\lambda = \lambda_1 + \lambda_2$. Hint: Start by computing $\mathbb{P}(Y > y)$. Two random variables with the same CDF have the same pdf. Why?

We start with computing $\mathbb{P}(Y > y)$, by substituting in the definition of Y .

$$\mathbb{P}(Y > y) = \mathbb{P}(\min(X_1, X_2) > y)$$

The probability that the minimum of two values is above a value is the probability that both of them are above that value. From there, we can separate them further because X_1 and X_2 are independent.

$$\mathbb{P}(X_1 > y \cap X_2 > y) = \mathbb{P}(X_1 > y)\mathbb{P}(X_2 > y) = e^{-\lambda_1 y}e^{-\lambda_2 y} = e^{-(\lambda_1 + \lambda_2)y} = e^{-\lambda y}$$

So $F_Y(y) = 1 - \mathbb{P}(Y > y) = 1 - e^{-\lambda y}$ and $f_Y(y) = \lambda e^{-\lambda y}$. Thus $Y \sim \text{Exponential}(\lambda)$, since this is the exact same CDF and PDF as an exponential distribution with parameter $\lambda = \lambda_1 + \lambda_2$.

- b) What is $\mathbb{P}(X_1 < X_2)$? Use the law of total probability. The law of total probability hasn't been covered in class yet, but will be soon at which point it would be good to revisit this problem!

By the continuous law of total probability,

$$\begin{aligned} \mathbb{P}(X_1 < X_2) &= \int_0^\infty \mathbb{P}(X_1 < X_2 \mid X_1 = x) f_{X_1}(x) dx \\ &= \int_0^\infty \mathbb{P}(X_2 > x) \lambda_1 e^{-\lambda_1 x} dx \\ &= \int_0^\infty e^{-\lambda_2 x} \lambda_1 e^{-\lambda_1 x} dx \\ &= \lambda_1 \int_0^\infty e^{-(\lambda_1 + \lambda_2)x} dx \\ &= \lambda_1 \left[\frac{-1}{\lambda_1 + \lambda_2} e^{-(\lambda_1 + \lambda_2)x} \right]_0^\infty \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2} \end{aligned}$$

- c) You have a digital camera that requires two batteries to operate. You purchase n batteries, labelled $1, 2, \dots, n$, each of which has a lifetime that is exponentially distributed with parameter λ , independently of all other batteries. Initially, you install batteries 1 and 2. Each time a battery fails, you replace it with the lowest-numbered unused battery. At the end of this process, you will be left with just one working battery. What is the expected total time until the end of the process? Justify your answer.

Let T be the time until the end of the process. We are trying to find $\mathbb{E}[T]$.

Let Y_i be the time we wait between the $(i - 1)$ -th battery failing and the i -th battery failing. Whenever we have two working batteries in the camera, we are waiting for the minimum of two independent $\text{Exponential}(\lambda)$ variables to fail. By part (a), the time until the first one fails is distributed as $Y_i \sim \text{Exponential}(2\lambda)$. Because the exponential distribution is memoryless, the battery that didn't fail is "as good as new", meaning the wait for the next failure is identically distributed, $Y_i \sim \text{Exponential}(2\lambda)$.

We start with 2 batteries and have $n - 2$ backups. The process ends when the n -th battery is installed and one of the two in the camera fails, leaving exactly 1 working battery. Thus, there are exactly $n - 1$ failure events that occur. The total time is

$T = Y_1 + Y_2 + \dots + Y_{n-1}$. By linearity of expectation:

$$\mathbb{E}[T] = \sum_{i=1}^{n-1} \mathbb{E}[Y_i] = \sum_{i=1}^{n-1} \frac{1}{2\lambda} = \frac{n-1}{2\lambda}$$

- d) In the scenario of the previous part, what is the probability that battery i is the last remaining battery as a function of i ? (You might want to use the memoryless property of the exponential distribution that has been discussed.)

If there are only two batteries i, j in the camera, by part (b), the probability each outlasts the other is $1/2$, since they have the same parameter. Hence, the last battery n has probability $1/2$ of being the last one remaining. The second to last battery $n-1$ has to beat out the previously remaining battery and the n -th battery, so the probability it lasts the longest is $(1/2)^2 = 1/4$. Working down inductively we get that the probability the i -th battery is the last remaining is $(1/2)^{n-i+1}$ for $i \geq 3$. Finally the first two batteries share the remaining probability as they start at the same time, with probability $(1/2)^{n-1}$ each.

11 Grading on a curve

In some classes (not CSE classes) an examination is regarded as being good (in the sense of determining a valid spread for those taking it) if the test scores of those taking it are well approximated by a normal density function. The instructor often uses the test scores to estimate the normal parameters μ and σ^2 and then assigns a letter grade of A to those whose test score is greater than $\mu + \sigma$, B to those whose score is between μ and $\mu + \sigma$, C to those whose score is between $\mu - \sigma$ and μ , D to those whose score is between $\mu - 2\sigma$ and $\mu - \sigma$ and F to those getting a score below $\mu - 2\sigma$. If the instructor does this and a student's grade on the test really is normally distributed with mean μ and variance σ^2 , what is the probability that student will get each of the possible grades A,B,C,D and F?

We can solve for each of these probabilities by standardizing the normal curve and then looking up each bound in the Z-table. Let X be the students score on the test. Then we have

$$\mathbb{P}(A) = \mathbb{P}(X \geq \mu + \sigma) = \mathbb{P}\left(\frac{X - \mu}{\sigma} \geq 1\right) = 1 - \mathbb{P}\left(\frac{X - \mu}{\sigma} < 1\right)$$

By the closure properties of the normal random variable, $\frac{X - \mu}{\sigma}$ is distributed as a standard normal random variable $Z \sim \mathcal{N}(0, 1)$. We can plug it into our Φ -table to get the following:

$$\mathbb{P}(A) = 1 - \Phi(1) = 1 - 0.84134 = 0.15866$$

The other probabilities can be found using a similar approach:

$$\mathbb{P}(B) = \mathbb{P}(\mu < X < \mu + \sigma) = \Phi(1) - \Phi(0) = 0.34134$$

$$\mathbb{P}(C) = \mathbb{P}(\mu - \sigma < X < \mu) = \Phi(0) - \Phi(-1) = 0.34134$$

$$\mathbb{P}(D) = \mathbb{P}(\mu - 2\sigma < X < \mu - \sigma) = \Phi(-1) - \Phi(-2) = 0.13591$$

$$\mathbb{P}(F) = \mathbb{P}(X < \mu - 2\sigma) = \Phi(-2) = 0.02275$$

12 Normal questions

- a) Let X be a normal random variable with parameters $\mu = 10$ and $\sigma^2 = 36$. Compute $\mathbb{P}(4 < X < 16)$.

Let $\frac{X-10}{6} = Z$. By the scale and shift properties of normal random variables $Z \sim \mathcal{N}(0, 1)$.

$$\begin{aligned}\mathbb{P}(4 < X < 16) &= \mathbb{P}\left(\frac{4-10}{6} < \frac{X-10}{6} < \frac{16-10}{6}\right) = \mathbb{P}(-1 < Z < 1) \\ &= \mathbb{P}(Z < 1) - \mathbb{P}(Z < -1) = \Phi(1) - \Phi(-1) = 0.68268\end{aligned}$$

- b) Let X be a normal random variable with mean 5. If $\mathbb{P}(X > 9) = 0.2$, approximately what is $\text{Var}(X)$?

Let $\sigma^2 = \text{Var}(X)$. Then,

$$\mathbb{P}(X > 9) = \mathbb{P}\left(\frac{X-5}{\sigma} > \frac{9-5}{\sigma}\right) = 1 - \mathbb{P}\left(\frac{X-5}{\sigma} < \frac{9-5}{\sigma}\right) = 1 - \Phi\left(\frac{4}{\sigma}\right) = 0.2$$

So, $\Phi\left(\frac{4}{\sigma}\right) = 0.8$. Looking up the phi values in reverse lets us undo the Φ function, and gives us $\frac{4}{\sigma} \approx 0.845$. Solving for σ we get $\sigma \approx 4.73$, which means that the variance is about 22.4.

- c) Let X be a normal random variable with mean 12 and variance 4. Find the value of c such that

$$\mathbb{P}(X > c) = 0.10.$$

$$\mathbb{P}(X > c) = \mathbb{P}\left(\frac{X-12}{2} > \frac{c-12}{2}\right) = 1 - \mathbb{P}\left(\frac{X-12}{2} < \frac{c-12}{2}\right) = 1 - \Phi\left(\frac{c-12}{2}\right) = 0.1$$

So, $\Phi\left(\frac{c-12}{2}\right) = 0.9$. Looking up the phi values in reverse lets us undo the Φ function, and gives us $\frac{c-12}{2} \approx 1.28$. Solving for c we get $c \approx 14.56$.

13 Do it in Reverse

- a) Let X be a normal random variable with parameters $\mu = 8$ and $\sigma^2 = 9$. Find x such that $\mathbb{P}(X \leq x) = 0.6$.

Let $\frac{X-8}{3} = Z$. By the scale and shift properties of normal random variables, $Z \sim \mathcal{N}(0, 1)$. Thus, we must find z such that $P(Z \leq z) = 0.6$.

$$\Phi(z) = P(Z \leq z) = 0.6$$

$$\Phi^{-1}(\Phi(z)) = \Phi^{-1}(0.6)$$

Thus, $z \approx 0.25$ by looking up the phi values in reverse to undo the Φ function. Then $\frac{x-8}{3} = z \approx 0.25$, so $x \approx 8.75$.

- b) Lots of statistics (like standardized test scores or heights) use *percentiles* to give context to where outcomes fall in a distribution. The n th percentile marks the outcome at which $n\%$ of the data points are less than the outcome. Let Y be a normal random variable with parameters $\mu = 15$ and $\sigma^2 = 4$. What value y marks the 85th percentile? What value b marks the 15th percentile?

We first find y , which marks the 85th percentile, so $\mathbb{P}(Y \leq y) = 0.85$. Let $\frac{Y-15}{2} = Z$. By the scale and shift properties of normal random variables, $Z \sim \mathcal{N}(0, 1)$. Thus, we must find z such that $P(Z \leq z) = 0.85$.

$$\Phi(z) = P(Z \leq z) = 0.85$$

$$\Phi^{-1}(\Phi(z)) = \Phi^{-1}(0.85)$$

Thus, $z \approx 1.04$ by looking up the phi values in reverse to undo the Φ function. Then $\frac{y-15}{2} = z \approx 1.04$, so $y \approx 17.08$.

Recall that normal distributions are symmetric around the mean, where $\mathbb{P}(Y \leq \mu) = 0.5$. Since

$$|\mathbb{P}(Y \leq \mu) - \mathbb{P}(Y \leq y)| = |0.5 - 0.85| = 0.35 = |\mathbb{P}(Y \leq \mu) - \mathbb{P}(Y \leq b)|,$$

$$b = \mu - |y - \mu| = 15 - |17.08 - 15| = 12.92,$$

so $b \approx 12.92$.

14 Bad Computer

Each day, the probability your computer crashes is 10%, independent of every other day. Suppose we want to evaluate the computer's performance over the next 100 days.

- a) Let X be the number of crash-free days in the next 100 days. What distribution does X have? Identify $\mathbb{E}[X]$ and $\text{Var}(X)$ as well. Write an exact (possibly unsimplified) expression for $\mathbb{P}(X \geq 87)$.

Since X counts the number of crash-free days (successes) in 100 days (trials), where each trial is a success with probability 0.9, we can see that X is binomial with $n = 100$ and $p = 0.9$, or $X \sim \text{Binomial}(100, 0.9)$. Hence, $\mathbb{E}[X] = np = 90$ and $\text{Var}(X) = np(1-p) = 9$. Finally,

$$\mathbb{P}(X \geq 87) = \sum_{k=87}^{100} \binom{100}{k} (0.9)^k (1 - 0.9)^{100-k}$$

- b) Approximate the probability of at least 87 crash-free days out of the next 100 days using the Central Limit Theorem. Use continuity correction.

Important: continuity correction says that if we are using the normal distribution to approximate

$$\mathbb{P}\left(a \leq \sum_{i=1}^n X_i \leq b\right)$$

where $a \leq b$ are integers and the X_i 's are i.i.d. **discrete** random variables taking all integer values, then, as our approximation, we should use

$$\mathbb{P}(a - 0.5 \leq Y \leq b + 0.5)$$

where Y is the appropriate normal distribution that $\sum_{i=1}^n X_i$ converges to by the Central Limit Theorem. The intuition here is that, to avoid a mismatch between discrete distributions (whose range is a set of integers) and continuous distributions, we get a better approximation by imagining that a discrete random variable, say W , is a continuous distribution with density function

$$f_W(x) := p_W(i) \quad \text{when } i - 0.5 \leq x < i + 0.5 \text{ and } i \text{ integer}$$

For more details see pages 209-210 in the Tsun book.

From the previous part, we know that $\mathbb{E}[X] = 90$ and $\text{Var}(X) = 9$.

$$\begin{aligned} \mathbb{P}(X \geq 87) &= \mathbb{P}(86.5 < X < 100.5) = \mathbb{P}\left(\frac{86.5 - 90}{3} < \frac{X - 90}{3} < \frac{100.5 - 90}{3}\right) \\ &\approx \mathbb{P}(-1.17 < Z < 3.5) \approx \Phi(3.5) - \Phi(-1.17) \\ &= \Phi(3.5) - (1 - \Phi(1.17)) \\ &= \Phi(3.5) + \Phi(1.17) - 1 \\ &\approx 0.9998 + 0.8790 - 1 = 0.8788 \end{aligned}$$

Notice that, if you had used $86.5 < X$ in place of $86.5 < X < 100.5$, your answer would have been nearly the same, because $\Phi(3.5)$ is so close to 1.

15 Another continuous r.v.

The density function of X is given by

$$f(x) = \begin{cases} a + bx^2 & \text{when } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

If $\mathbb{E}[X] = \frac{3}{5}$, find a and b .

To find the value of two variables, we need two equations to solve as a system. We know that $\mathbb{E}[X] = \frac{3}{5}$, so we know, by the definition of expected value, that

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx = \frac{3}{5}$$

Since $f(x)$ is defined to be 0 outside of the given range, we can integrate within only that range, plugging in $f(x)$:

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^0 xf(x)dx + \int_0^1 xf(x)dx + \int_1^{\infty} xf(x)dx = \int_0^1 x(a + bx^2)dx \\ &= \int_0^1 (ax + bx^3)dx = \left(\frac{ax^2}{2} + \frac{bx^4}{4} \right) \Big|_0^1 = \frac{a}{2} + \frac{b}{4} = \frac{3}{5}\end{aligned}$$

We also know that a valid density function integrates to 1 over all possible values. Thus, we can perform the same process to get a second equation:

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^0 f(x)dx + \int_0^1 f(x)dx + \int_1^{\infty} f(x)dx \\ &= \int_0^1 (a + bx^2)dx = \left(ax + \frac{bx^3}{3} \right) \Big|_0^1 = a + \frac{b}{3} = 1\end{aligned}$$

Solving this system of equations we get that $a = \frac{3}{5}, b = \frac{6}{5}$

16 Point on a line

A point is chosen at random on a line segment of length L . Interpret this statement and find the probability that the ratio of the shorter to the longer segment is less than $\frac{1}{4}$.

Define RV X to be the distance of your random point from the leftmost side of the stick. Since we're choosing a point at random, this RV has an equal likelihood of any distance from 0 to L , making it a continuous uniform RV with parameters $a = 0, b = L$. For the ratio to be less than $\frac{1}{4}$, the shorter segment has to be less than $\frac{L}{5}$ in length.

This can happen when $X < \frac{L}{5}$ or $X > \frac{4L}{5}$. Thus, using the CDF of a continuous uniform distribution (and the fact that $\mathbb{P}(X = k) = 0$ for any k since X is a continuous random variable), the probability that the ratio is less than $\frac{1}{4}$ is

$$\mathbb{P}(X \leq \frac{L}{5}) + \mathbb{P}(X > \frac{4L}{5}) = F_X(\frac{L}{5}) + (1 - F_X(\frac{4L}{5})) = \frac{\frac{L}{5} - 0}{L - 0} + (1 - \frac{\frac{4L}{5} - 0}{L - 0}) = \frac{1}{5} + (1 - \frac{4}{5}) = \frac{2}{5}$$

17 Transforming continuous random variables

The next few questions are designed to help you understand the issue with transforming continuous random variables.

Specifically let's explore why we cannot simply adapt the discrete formula $p_Y(y) = \sum_{x|g(x)=y} p_X(x)$ into an integral $f_Y(y) = \int_{x|g(x)=y} f_X(x) dx$ for continuous random variables.

- a) Suppose $X \sim \text{Unif}(-1, 1)$ and $Y = X^2$. We want to find the density of Y at $y = 0.25$. If we blindly apply the incorrect formula $f_Y(0.25) = \int_{x|x^2=0.25} f_X(x) dx$, what is the mathematical result of this specific integral?
- (a) $f_X(-0.5) + f_X(0.5) = 1$
 - (b) 0
 - (c) 0.5
 - (d) 0.25

Correct: (b)

The set of points where $x^2 = 0.25$ is exactly two discrete points: $\{-0.5, 0.5\}$. In calculus, the integral of any finite continuous function over a discrete set of points (a set of “measure zero”) is always 0.

This shows why the formula is mathematically broken: you cannot integrate over a set of distinct points to get a non-zero density!

- b) Suppose $X \sim \text{Unif}(0, 1)$ and we apply the transformation $Y = 3X$. We know that Y is uniformly distributed over $(0, 3)$, so its true density should be $f_Y(y) = 1/3$ for $y \in (0, 3)$. If a student mistakenly assumes they can just “move” the density from X to Y by setting $f_Y(y) = f_X(x)$ where $3x = y$, what incorrect density would they get for Y ?
- (a) $f_Y(y) = 1$ for $y \in (0, 3)$
 - (b) $f_Y(y) = 3$ for $y \in (0, 3)$
 - (c) $f_Y(y) = 1/3$ for $y \in (0, 1)$
 - (d) $f_Y(y) = 0$ for $y \in (0, 3)$

Correct: (a)

For any $y \in (0, 3)$, the corresponding x is $y/3$. Because $y/3$ is between 0 and 1, the density of X there is $f_X(y/3) = 1$ using the probability density function for a continuous uniform random variable.

If the student sets $f_Y(y) = 1$ over the interval $(0, 3)$, the total area under their PDF would be $\int_0^3 1 dy = 3$, which violates the rule that all PDFs must integrate to 1. This error happens because the transformation $Y = 3X$ “stretches” the variable by a factor of 3, so the density must scale down by a factor of $1/3$ to conserve the total probability mass of 1.

- c) Based on the previous examples, why does using the CDF method ($F_Y(y) = \mathbb{P}[Y \leq y]$) succeed where manipulating the PDF directly fails?
- (a) The CDF converts the continuous variable into a discrete variable before taking the derivative.
 - (b) The CDF works over intervals (areas) rather than individual points, properly accounting for how the transformation stretches or squishes the probability mass.
 - (c) The CDF method only works for linear functions, bypassing the need for integration.
 - (d) Probability density functions cannot be evaluated at specific points; they are only defined at infinity.

Correct: (b)

Continuous probability is fundamentally about the area over an interval, not the height at a specific point. By starting with $F_Y(y) = \mathbb{P}[Y \leq y]$, we are looking at an interval. When we map that interval back to X and eventually take the derivative, the chain rule automatically generates the scaling factor (the Jacobian) that accounts for how the transformation stretched or squished the geometry of the sample space.

18 Non-Monotonic Transformations

When a transformation is not strictly increasing or decreasing, we have to be extra careful when setting up the inequalities for the CDF. Let $X \sim \text{Unif}(-2, 2)$ and let $Y = X^2$.

- (a) What is the cumulative distribution function (CDF) of Y ? Be sure to clearly state the range of possible values for Y .

First, we determine the support of Y . Since X ranges from -2 to 2 , the square of X will range from 0 to 4 . Therefore, the support of Y is $[0, 4]$.

For $y \in [0, 4]$, we find the CDF by setting up the probability:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y)$$

Taking the square root of both sides of an inequality introduces both a positive and a negative bound:

$$F_Y(y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y})$$

We can rewrite this probability in terms of the CDF of X :

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

For $X \sim \text{Unif}(-2, 2)$, the CDF on its support is $F_X(x) = \frac{x - (-2)}{2 - (-2)} = \frac{x+2}{4}$. Substituting our bounds into this formula gives:

$$F_Y(y) = \left(\frac{\sqrt{y} + 2}{4} \right) - \left(\frac{-\sqrt{y} + 2}{4} \right) = \frac{2\sqrt{y}}{4} = \frac{\sqrt{y}}{2}$$

The complete piecewise CDF is:

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \frac{\sqrt{y}}{2} & 0 \leq y \leq 4 \\ 1 & y > 4 \end{cases}$$

- (b) Derive the probability density function (pdf) of Y by taking the derivative of your answer from part (a).

To find the pdf, we take the derivative of the CDF with respect to y over the valid support range $(0, 4]$:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \left(\frac{1}{2} y^{1/2} \right)$$

Therefore,

$$f_Y(y) = \frac{1}{2} \cdot \frac{1}{2} y^{-1/2} = \frac{1}{4\sqrt{y}}.$$

The complete piecewise PDF is:

$$f_Y(y) = \begin{cases} \frac{1}{4\sqrt{y}} & 0 < y \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

Note that the pdf goes to infinity as y approaches 0, which is a common feature when transforming continuous distributions at stationary points (like the vertex of $y = x^2$).

19 Transformations

Suppose $X \sim \text{Uniform}(0, 1)$ has the continuous uniform distribution on $(0, 1)$. Let $Y = -\frac{1}{\lambda} \ln X$ for some $\lambda > 0$.

- a) What is Ω_Y ?

$\Omega_Y = (0, \infty)$ because $\ln(x) \in (-\infty, 0)$ for $x \in (0, 1)$. Thus, that range times a necessarily negative number $-\frac{1}{\lambda}$, will result in a range from 0 to positive infinity.

- b) First write down $F_X(x)$ for $x \in (0, 1)$. Then, find $F_Y(y)$ on Ω_Y .

$F_X(x) = x$ for $x \in (0, 1)$ because that is the CDF of the continuous uniform distribution. We find the CDF of Y by plugging in the given definition of Y and getting into a form where we can use the CDF of X . Let $y \in \Omega_Y$.

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}\left(-\frac{1}{\lambda} \ln X \leq y\right) = \mathbb{P}(\ln X \geq -\lambda y) = \mathbb{P}(X \geq e^{-\lambda y}) = 1 - \mathbb{P}(X < e^{-\lambda y})$$

Then, because $e^{-\lambda y} \in (0, 1)$:

$$= 1 - F_X(e^{-\lambda y}) = 1 - e^{-\lambda y}$$

- c) Now find $f_Y(y)$ on Ω_Y (by differentiating $F_Y(y)$ with respect to y). What distribution does Y have?

$$f_Y(y) = F'_Y(y) = \lambda e^{-\lambda y}$$

Hence, $Y \sim \text{Exponential}(\lambda)$.

20 Tweets

A prolific twitter user tweets approximately 350 tweets per week. Let's assume for simplicity that the tweets are independent, and each consists of a uniformly random number of characters between 10 and 140. (Note that this is a discrete uniform distribution.) Thus, the central limit theorem (CLT) implies that the number of characters tweeted by this user is approximately normal with an appropriate mean and variance. Assuming this normal approximation is correct, estimate the probability that this user tweets between 26,000 and 27,000 characters in a particular week. (This is a case where continuity correction will make virtually no difference in the answer, but you should still use it to get into the practice!).

Let X be the total number of characters tweeted by a twitter user in a week. Let $X_i \sim \text{Unif}(10, 140)$ be the number of characters in the i th tweet (since the start of the week). Since X is the sum of 350 i.i.d. rvs with mean $\mu = 75$ and variance $\sigma^2 = 1430$, $X \approx N \sim \mathcal{N}(350 \cdot 75, 350 \cdot 1430)$. Thus,

$$\begin{aligned} \mathbb{P}(26,000 \leq X \leq 27,000) &= \mathbb{P}(25,999.5 \leq X \leq 27,000.5) \\ &\approx \mathbb{P}(25,999.5 \leq N \leq 27,000.5) \end{aligned}$$

Standardizing this gives the following formula

$$\begin{aligned} \mathbb{P}(25,999.5 \leq N \leq 27,000.5) &= \mathbb{P}\left(\frac{25,999.5 - 350 \cdot 75}{\sqrt{350 \cdot 1430}} \leq \frac{N - 350 \cdot 75}{\sqrt{350 \cdot 1430}} \leq \frac{27,000.5 - 350 \cdot 75}{\sqrt{350 \cdot 1430}}\right) \\ &\approx \mathbb{P}\left(-0.35 \leq \frac{N - 350 \cdot 75}{\sqrt{350 \cdot 1430}} \leq 1.06\right) \\ &\approx \mathbb{P}(-0.35 \leq Z \leq 1.06) \\ &= \Phi(1.06) - \Phi(-0.35) \\ &\approx 0.85543 - (1 - 0.63683) \\ &= 0.49226 \end{aligned}$$

So the probability that this user tweets between 26,000 and 27,000 characters in a particular week is approximately 0.4923.

21 Ping Pong

You're playing ping pong with your friend, and want to keep playing until you've scored 15 points. Unfortunately, your friend is a much more skilled ping pong player than you, so you only win points 25% of the time (with each point being independent of the other points). Approximate the probability that you'll need to play at least 50 points before stopping.

Let X be the total number of points played. We want to approximate $\mathbb{P}(X \geq 50)$.

Let X_i be the number of points played starting after the $i - 1$ th point you win and up to and including the i th point you win, with X_1 the number of points up to and including the first point you win. Then, we have $X = \sum_{i=1}^{15} X_i$. Because you win each point independently with probability 0.25, we have $X_i \sim \text{Geo}(0.25)$. Thus,

$$\mathbb{E}[X_i] = 4$$

and

$$\text{Var}(X_i) = \frac{1 - 0.25}{(0.25)^2} = 12$$

Since X is the sum of 15 i.i.d. r.v.s with mean $\mu = 4$ and variance $\sigma^2 = 12$, by the central limit theorem we have $X \approx N \sim \mathcal{N}(15 \cdot 4, 15 \cdot 12)$. Thus,

$$\begin{aligned} \mathbb{P}(X \geq 50) &= \mathbb{P}(X \geq 49.5) && \text{Continuity correction} \\ &\approx \mathbb{P}(N \geq 49.5) && \text{CLT} \end{aligned}$$

Standardizing, we get the following:

$$\begin{aligned} \mathbb{P}(N \geq 49.5) &= \mathbb{P}\left(\frac{N - 15 \cdot 4}{\sqrt{15 \cdot 12}} \geq \frac{49.5 - 15 \cdot 4}{\sqrt{15 \cdot 12}}\right) \\ &\approx \mathbb{P}\left(\frac{N - 15 \cdot 4}{\sqrt{15 \cdot 12}} \geq -0.782\right) \\ &\approx \mathbb{P}(Z \geq -0.782) \\ &= 1 - \mathbb{P}(Z \leq -0.782) \\ &= 1 - \Phi(-0.782) \\ &= 1 - (1 - \Phi(0.782)) \\ &= \Phi(0.782) \\ &\approx 0.7823 \end{aligned}$$