STA 131A: Introduction to Probability Theory

Lecture 19: Derived Distributions (cont'd) + Review for Midterm 2

Dogyoon Song

Spring 2026, UC Davis

Announcement

Midterm 2 on Fri, May 15 (10:00 am–10:50 am in class)

- **Arrive early**: The exam starts at 10:00 am and ends at 10:50 am sharp
- **One hand-written cheat sheet:** Letter-size (8.5"×11"), double-sided, brief formulas/notes
- **Calculator**: A simple (non-graphing) scientific calculator is allowed
- **No other materials** beyond the single cheat sheet (no textbooks, etc.) is allowed
- **SDC accommodations:** Confirm scheduling with AES online ASAP

Discussion section tomorrow (Thu, May 14)

- The TA will go over two practice midterms

Office hour today (Wed, May 13) 2:30–3:30 PM

Agenda

Last time: Derived distributions

- Functions of one random variable: $Z = g(X)$
  - Generic method: support $\to$ CDF $\to$ PDF
  - Monotone transformation formula
  - Linear transformations as a special case
- Functions of two random variables: $Z = g(X, Y)$
  - CDF by integrating over a region

Today: Derived distributions, continued + Midterm 2 review

- Finish functions of two random variables
  - CDF-first method for $Z = h(X, Y)$
  - Convolution for sums of independent random variables
- Brief review for Midterm 2
  - Continuous PDFs/CDFs, normal variables, joint PDFs
  - Conditioning, Bayes, conditional expectation
  - Derived distributions

Recap: Derived distributions

Generic, CDF-first workflow for $Z = g(X)$

1. Identify the support: possible values of $Z = g(X)$.
2. Compute the CDF:
$$F_Z(z) = P(Z \le z) = P(g(X) \le z).$$
3. Differentiate on the support:
$$f_Z(z) = \frac{d}{dz} F_Z(z).$$
4. Set $f_Z(z) = 0$ outside the support.

PDF formula for a strictly monotone transformation

Suppose X is continuous and $Z = g(X)$, where g is strictly monotone and differentiable on the support of X, with inverse
$$x = h(z) = g^{-1}(z).$$
Then, for z in the support of Z,
$$f_Z(z) = f_X\big(h(z)\big) \cdot \left|h'(z)\right|.$$

Recap: Function of two random variables

Let (X, Y) be jointly continuous, and let

$$Z = g(X, Y).$$

To find the distribution of Z, a general strategy is again to identify the CDF first:

$$F_Z(z) = P(Z \leq z) = P(g(X, Y) \leq z) = \iint_{\{(x,y):g(x,y)\leq z\}} f_{X,Y}(x, y)\, dx\, dy.$$

Then, when differentiable,

$$f_Z(z) = F_Z'(z).$$

Main challenge: Identifying the integration region correctly.

Practical rules: Draw the region $\{(x, y) : g(x, y) \leq z\}$. Split the integral as needed whenever the boundary changes.

- Recall the example: Let (X, Y) be uniformly distributed on the unit square $[0, 1]^2$, and let

$$Z = XY.$$

Sum of independent random variables: Convolution

Definition (Convolution)

Let (X, Y) be jointly continuous and let $Z = X + Y$.
If X and Y are independent, the density of $Z = X + Y$ is obtained by the **convolution** operation:

$$f_Z(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z - x)\, dx.$$

The integral only runs over values of x for which both x and $z - x$ lie in the supports of X and Y.

- **Proof idea:** $f_{X,Z}(x, z) = f_{X,Y}(x, z - x) = f_X(x) f_Y(z - x)$ by independence
- **Support rule:** The integral only includes x's for which both

$$x \in \text{supp}(X) \qquad \text{and} \qquad z - x \in \text{supp}(Y).$$

- The convolution of PMFs is defined similarly: $p_Z(z) = (p_X * p_Y)(z) = \sum_x p_X(x)\, p_Y(z - x)$.
- **Interpretation:** To get $X + Y = z$, sum over all possible ways $x + (z - x) = z$.

Example: Sum of two independent uniforms

Example

Let $X, Y \overset{\text{independent}}{\sim} \mathrm{Uniform}(0,1)$, and $Z = X + Y$.

Question: Find the PDF of Z.

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx.$$

Since $f_X(x) = 1$ for $0 < x < 1$ and $f_Y(z-x) = 1$ when $0 < z - x < 1$, the integrand $f_X(x)f_Y(z-x)$ is nonzero only when

$$0 < x < 1 \qquad \text{and} \qquad z - 1 < x < z.$$

Thus x must lie in the interval

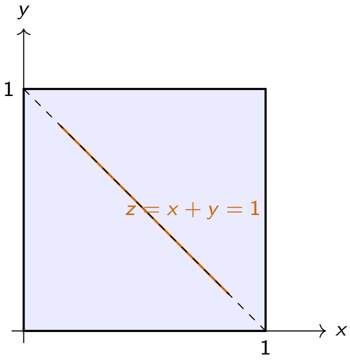
$$\max(0, z-1) < x < \min(1, z).$$

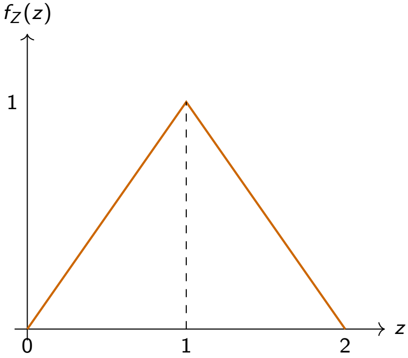
The density of Z at z is equal to the length of this interval (whenever it is positive):

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

Geometric view of $X + Y$

For $X, Y \sim \mathrm{Uniform}(0, 1)$, the joint density is uniform on the unit square.





For a thin band $z < X + Y < z + \Delta z$, the probability is the area of the band. As $\Delta z \to 0$, this area is proportional to the length of the slice $x + y = z$ inside the unit square.

This length increases for $0 < z < 1$ and decreases for $1 < z < 2$, producing a triangular density.

Example: Sum of independent exponentials

Example

Let X, Y be independent $\mathrm{Exponential}(\lambda)$ random variables, and let

$$Z = X + Y.$$

Question: Find the PDF of Z.

Since $X, Y > 0$, we have $Z > 0$. Using convolution, for $z > 0$,

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx = \int_0^z \lambda e^{-\lambda x} \lambda e^{-\lambda(z-x)}\, dx \\ &= \lambda^2 e^{-\lambda z} \int_0^z dx = \lambda^2 z e^{-\lambda z}. \end{aligned}$$

Thus,

$$f_Z(z) = \begin{cases} \lambda^2 z e^{-\lambda z}, & z > 0, \\ 0, & z \le 0. \end{cases}$$

Remark: This distribution is often called an Erlang/Gamma distribution with shape 2.

Pop-up quiz: Convolution

Let X, Y be independent $\mathrm{Uniform}(0,1)$, and let

$$Z = X + Y.$$

Which describes the density of Z correctly?

A) $f_Z(z) = \frac{1}{2}, \quad 0 < z < 2$

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

C) $f_Z(z) = 1, \quad 0 < z < 2$

D) $f_Z(z) = z, \quad 0 < z < 2$

Answer: B.

For fixed z, the contributing values of x satisfy

$$\max(0, z-1) < x < \min(1, z).$$

The length is z for $0 < z < 1$, and $2 - z$ for $1 < z < 2$.

Midterm 2 review: What to know cold

Core skills

- Work with PDFs and CDFs:

$$P(a < X \le b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\,dx.$$

- Standardize normal random variables:

$$Z = \frac{X - \mu}{\sigma}.$$

- Compute marginals and conditional densities from joint PDFs.
- Use conditional expectation and total expectation.
- Derive distributions using CDF-first, inverse transformations, or convolution.

Main exam advice: identify the support and the relevant region before integrating.

Review: Continuous random variables

Continuous random variable and PDF

$$P(X \in A) = \int_A f_X(x)\,dx, \qquad f_X(x) \geq 0, \qquad \int_{-\infty}^{\infty} f_X(x)\,dx = 1.$$

CDF

$$F_X(x) = P(X \leq x), \qquad P(a < X \leq b) = F_X(b) - F_X(a).$$

If X is continuous and F_X is differentiable, then

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt, \qquad f_X(x) = F_X'(x).$$

Warning: a density value is not a probability; probabilities come from integration.

Review: Normal random variables

Normal random variable

$$X \sim N(\mu, \sigma^2)$$

means X has mean μ and variance σ^2.

Standardization

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Thus,

$$P(a < X \leq b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

Common mistake: in $N(\mu, \sigma^2)$, σ is the standard deviation and σ^2 is the variance.

Review: Joint PDFs and marginalization

Joint PDF

$$P((X, Y) \in A) = \iint_A f_{X,Y}(x, y)\, dx\, dy.$$

Marginals

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx.$$

Expectation

$$\mathbb{E}[g(X, Y)] = \iint g(x, y) f_{X,Y}(x, y)\, dx\, dy.$$

Warning: integration limits come from the geometry of the support.

Review: Conditioning and independence

Conditional PDF

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Conditional expectation

$$\mathbb{E}[X \mid Y = y] = \int x \, f_{X|Y}(x \mid y) \, dx.$$

Total expectation

$$\mathbb{E}[X] = \int \mathbb{E}[X \mid Y = y] \, f_Y(y) \, dy.$$

Independence

$$f_{X,Y}(x,y) = f_X(x) f_Y(y), \qquad \forall (x,y).$$

Review: Bayes' rule

Continuous hidden state

$$f_{X|Y}(x \mid y) = \frac{f_X(x) f_{Y|X}(y \mid x)}{\int f_X(t) f_{Y|X}(y \mid t)\, dt}.$$

Discrete hidden state, continuous observation

$$P(H = h \mid Y = y) = \frac{P(H = h) f_{Y|H}(y \mid h)}{\sum_{h'} P(H = h') f_{Y|H}(y \mid h')}.$$

Implication:

$$\text{posterior} \propto \text{prior} \times \text{likelihood.}$$

Review: Derived distributions

Generic CDF-first method

$$F_Z(z) = P(g(X) \le z), \qquad f_Z(z) = F_Z'(z).$$

Strictly monotone formula If $Z = g(X)$, g is differentiable and strictly monotone, and $x = h(z) = g^{-1}(z)$, then

$$f_Z(z) = f_X(h(z))|h'(z)|.$$

Function of two variables

$$F_Z(z) = P(h(X, Y) \le z) = \iint_{\{(x,y):h(x,y)\le z\}} f_{X,Y}(x, y)\, dx\, dy.$$

Convolution If $Z = X + Y$ and X, Y are independent,

$$f_Z(z) = \int f_X(x) f_Y(z - x)\, dx.$$

Midterm 2 review: Common pitfalls

- A density value is not a probability; probabilities come from integration.
- For continuous variables, endpoints usually do not matter.
- In $N(\mu, \sigma^2)$, σ^2 is the variance and σ is the standard deviation.
- For joint densities, integration limits come from the geometry of the support.
- Independence requires factorization:

$$f_{X,Y}(x, y) = f_X(x) f_Y(y), \qquad \text{for all } (x, y).$$

- For transformations, support restrictions are part of the answer.

Wrap-up: How to prepare

Use the practice midterms for review

- First attempt them without notes and under time pressure.
- Mark which steps failed: identifying the tool, setting up the model, or doing the calculation.
- Then review and prepare your cheat sheet around those weak points.

When preparing your cheat sheet

- Consider including definitions and assumptions, not only formulas.
- Ensure you are fluent with the tools you have written.

During the exam

- Succinctly write down the relevant event or random variable before using a formula.
- Keep notation clear: distinguish events, values, PMFs, and expectations.
- If stuck, identify the support, draw the region, or condition on a useful variable/event.

Advice: Be fluent in translating words to probabilistic objects and choosing the right tool.